

**This Page Is Inserted by IFW Operations  
and is not a part of the Official Record**

## **BEST AVAILABLE IMAGES**

**Defective images within this document are accurate representations of the original documents submitted by the applicant.**

**Defects in the images may include (but are not limited to):**

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	DB	ID	Description
1	1298	83.2	248	22	AA041224	Human polypeptide
2	1255	80.4	235	22	AA039438	Human polypeptide
3	1255	80.4	235	23	AA047664	MOLa protein sequ
4	1255	80.4	235	23	AA047665	MOLb protein sequ
5	1250	80.1	235	23	AA076316	Human novel serine
6	1247	79.9	235	22	AA007697	Human Zt1p3 serin
7	1247	79.9	235	22	AA027666	Human protein APP8
8	1247	79.9	235	23	AA074746	Human protease PRP
9	1247	79.9	235	24	ABG73741	Human serine prote
10	1222	78.3	265	23	AA087744	Amino acid sequenc
11	1212	77.9	228	23	AA047674	Protein sequence o
12	352.5	22.6	247	21	AA021321	Human trypsinogen.
13	350.5	22.5	246	21	AA078975	Canine antionic tryp
14	349.5	22.4	224	20	AA011160	Human trypsin seriti
15	349.5	22.4	230	22	AA098503	Human trypsin seriti
16	349.5	22.4	225	22	AA098508	Human TRYI trypsin
17	347.5	22.3	241	21	AA021316	Human trypsinogen
18	346.5	22.2	247	19	AA064260	Human amyloid beta
19	346	22.2	246	9	AA081243	Human spleen tryps
20	346	22.2	247	16	AA082703	Human pancreatic t
21	346	22.2	247	21	AA078974	Canine cationic tryp
22	344.5	22.1	240	19	AA057740	Trypsinogen-like p
23	335.5	21.5	228	23	AA098259	Synthetic shortene
24	335.5	21.5	247	18	AA008475	Porcine trypsinogen
25	335.5	21.5	247	23	AA098258	Pig trypsinogen SE
26	334.5	21.4	247	23	AA087693	Human pancreatic t
27	334.5	21.4	261	24	AB041530	Human D1HP protei
28	328.5	21.1	247	21	AA035701	Human trypsin hl a
29	317.5	20.4	230	15	AA035638	Bovine trypsinogen
30	312.5	20.0	223	20	AA081767	Bovine TRYP peptid
31	312.5	20.0	223	21	AA069973	Bovine TRYP peptid
32	312.5	20.0	223	23	AB078122	Amino acid sequenc
33	312.5	20.0	223	23	AB031841	Example protein #3
34	312.5	20.0	223	23	AB083122	Partial trypsin se
35	312.5	20.0	224	15	AA056367	Bovine trypsin. B
36	312.5	20.0	229	21	AA077494	Bovine trypsinogen
37	312.5	20.0	231	21	AA091926	Recombinant trypsi
38	312.5	20.0	231	22	AA080953	Bovine met-phe-try
39	312.5	20.0	233	21	AA091925	Trypsinogen analog
40	311	19.9	243	22	AA055891	Amino acid sequenc
41	305	19.6	241	22	AA055890	Amino acid sequenc
42	305	19.6	241	22	AA012295	Human PRO polypept
43	305	19.6	241	24	AB066693	Human PRO polypept
44	305	19.6	241	24	AB066696	Human secreted/tra
45	305	19.6	241	24	AB057774	Novel secreted and

RESULT 1  
AA041224  
ID AA041224 standard; Protein; 248 AA.  
XX  
XX AA041224;  
XX  
XX 22-OCT-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 6155.  
XX  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoclastic;  
KW chemokilnetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukemia.  
XX  
XX Homo sapiens.  
XX

XX MO200153312-A1.  
 PN  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000MO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI, 2001-442253/47.  
 DR N-PSDB; AAI60380.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 2; SEQ ID NO 6155; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemoclastic/chemokine activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 CC  
 XX  
 XX  
 SQ Sequence 248 AA;  
 Alignment Scores:  
 Pred. No.: 3,53e-131 Length: 248  
 Score: 1298.00 Matches: 244  
 Percent Similarity: 98.39% Conservative: 0  
 Best Local Similarity: 98.39% Mismatches: 4  
 Query Match: 83.21% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-037-270-482 (1-866) x AAM41224 (1-248)  
 QY 3 CACTTACTCCCTGAGTAAAGGGGAGAGCTGATCACCATGAATATGCTTCTATTGG 62  
 DB 1 HisleuLeuProGluLeuAArgGlyLysSerTriplethetLeuTyrValPheTyrLeu 20  
 QY 63 GGATGCTCGCTGGGACATTTTCTTGTGCTGACTCATCTGTTCAGAAAGAACCTGCT 122  
 DB 21 GlyValLeuAacGlyThrPhePhePheAlaAspSerSerValGlnLysGluAspProAla 40  
 QY 123 CCTATTGTTGTAACCCAGATGTCATCAACCCCGTGGCGGCTCATCAAAACC 182  
 DB 41 ProTyrLeuValTyrLeuLysSerHisPheAsnProCysValGlyValLeuLysPro 60  
 QY 183 AGCTGGGTGCTGGCCCGAGCTCACTGCTATTCAAAATGGAAGTATGCTGGGAAT 242

DB 61 SerTriPValLeuAlaProAlaHisCysTyrLeuProAsnLeuLysValMetLeuGlyAsn 80  
 QY 243 TTCAGAGAGAGAGTACAGAGCGGTACTGAAAGACAAATTAAACCCATTGATGTCGCG 302  
 DB 81 PheLysSerArgValAlaArgAspGlyLynGluGlnThrIleAsnProIleGlnIleValArg 100  
 QY 303 TACTGAACTACAGTCATAGCGCCCAAGAGATGATGCTCATCAAGCTGAGCTAAG 362  
 DB 101 TyrTriPAsnTyrSerHisSerIleAlaProGlnAspAspLeuMetLeuLysLeuAlaLys 120  
 QY 363 CCTGCCATGCTCATATCCCAAGTCCAGCCCTTCCCTGCCACCAATGTCAGGCCA 422  
 DB 121 ProAlaMetLeuAsnProLysValGlnAlaIleAsnProProThrThrAsnValArgPro 140  
 QY 423 GGCACGTGCTGTACTCTCAGGTTTGATCGTGGAGCCAGAAAGAGTGGCGACACCT 482  
 DB 141 GlyThrValCysLeuLeuSerGlyLeuAspTriPSerGlnGluAsnSerGlyArgHisPro 160  
 QY 483 GACTTCGCGCAGAACTGAGAGCGCCCGCTGATGTCATCGAGAAATGCCAAAAACAGAA 542  
 DB 161 AspLeuAArgGlnAsnLeuGlnAlaProValMetSerAspArgGlyCysGlnLysThrGlu 180  
 QY 543 CAAAGAAAAAGCCACAGAAATTCCTTATGTGTGAAATTTGTGAAAGTATTACGCCAAT 602  
 DB 181 GlnGlyLysSerHisArgAsnSerLeuCysValLysPheValLysPheSerArgIle 200  
 QY 603 TTGGGAGAGTGGCGCTGCTGCTGATCTGCAAGAAAGAGCTCCAGGGAATCGAGGTG 662  
 DB 201 PheGlyGluValAlaValAlaThrValIleCysLysAspLysLeuGlnLysIleGluVal 220  
 QY 663 GGGCACTTCATGAGGAGGAGCGTGGCATCTACCAATGTTTCAAAATATGATCTG 722  
 DB 221 GlyHisPheMetCglyLysPheValGlyIleTyrThrAsnValTyrLysTyrValSerTriP 240  
 QY 723 ATTGAGAACATGCTTAAGACAG 746  
 DB 241 IleGluAsnThrAlaLysAspLys 248  
 RESULT 2  
 AAM39438  
 ID AAM39438 standard; Protein; 235 AA.  
 XX  
 AC AAM39438;  
 XX  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 2583.  
 XX  
 KM Human; neurotropic; immunosuppressant; cyostatic; gene therapy; cancer;  
 KM peripheral nervous system; neuropathy; central nervous system; CNS;  
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoclastic;  
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KM leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000MO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.



XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J,  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AA158594.  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 4; SEQ ID NO 2583; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AA38642-AA42213) with neurotropic,  
 CC immunosuppressant and cyrostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Actvlin/inhibin activity, chemoclastic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 XX  
 SQ Sequence 235 AA;  
 Alignment Scores:  
 Pred. No.: 1.53e-126 Length: 235  
 Score: 1255.00 Matches: 235  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 80.45% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-037-270-482 (1-866) x AAM39438 (1-235)  
 QY 42 ATGAATATGCTCTTATTTGGTGTCTTCCTGGGACATTTTCTTGTGACTCATCT 101  
 DB 1 MetLysTyrValIlePheTyrLeuGlyValLeuAlaGlyThrPhePheAlaAspSer 20  
 QY 102 GTTCAGAAAGACACCTGCTCTCTATTTGGTGTACCTCAAGTCTCACTTCAACCCCTGT 161  
 DB 21 ValGlnLysGlnAspProAlaProTyrLeuValTyrLeuLysSerHisPheAsnProCys 40  
 QY 162 GTGGGGCTCTCATCAAAACCACTGGGTGGTGGCCCACTCACTGCTCTTTTACCAAT 221  
 DB 41 ValGlyValLeuLysProSerTrpValLeuAlaProAlaHisCysTyrLeuProAsn 60  
 QY 222 CTGAAGTGTGCTGGGAAATTTCAAGACAGAGTGTGAGAGCGTACTGAACAGACAAAT 281  
 DB 61 LeuLysValMetLeuLysPheAsnLysSerArgValArgAspLysThrGlnThrIle 80  
 QY 282 AACCCATTCAGATGCTCGCTACTGTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 341  
 DB 81 AsnProLysGlnLysValArgTyrTrpAsnTyrSerHisSerLysProGlnAspLys 100  
 QY 342 ATGCTCATCAAGTGTGCTAAGCTGCGCATCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
 DB 101 MetLeuLysValLeuLysPheAlaMetLeuAsnProLysValGlnProLeuProLeu 120  
 QY 402 GCCACCAACATGCTGAGCGGACGAGGACTGTGTCTACTCTCAGGTTTGAAGGACAA 461  
 DB 121 AlaThrThrAsnValArgProGlyThrValCysLeuLysSerLysLeuAspTrpSerGln 140  
 QY 462 GAAACAGTGGCGGACACCTGACTTGGCGGACAGAACTGTGAGGCCCTGATGTCTGAT 521  
 DB 141 GluAsnSerCysLysArgHisProAspLeuArgGlnAsnLeuLysValProValMetSerAsp 160

QY 522 CGAATGCCAAAAACAGAACAGAAAAACACAGAAATTCCTATGTGAAATT 581  
 DB 161 ArgGluCysGlnLysPheThrLysGlnLysSerHisArgAsnSerLeuValLysPhe 180  
 QY 582 GTGAAGATTCAGCCGAATTTTGGGAGGTGGCCGTGCTACTGCTCATTCGCAAGAC 641  
 DB 181 ValLysValPheSerArgLysPheGlyValAlaValAlaThrValLysCysLysAsp 200  
 QY 642 AAGCTCCAGGAATGAGGTGGGCACTTCATGCGAGGGAGAGTGGCATCTTACCAAT 701  
 DB 201 LysLeuGlnGlyLysGlnValGlyHisPheMetGlyAspValGlyLysThrAsn 220  
 QY 702 GTTACAATATGATCTCTGATGAGAACACTGTCAAGACACAG 746  
 DB 221 ValTyrLysTyrValSerTrpLysGlnAsnThrAlaLysAspLys 235  
 RESULT 3  
 AAM47664  
 ID AAM47664 standard; Protein; 235 AA.  
 AC  
 XX AAM47664;  
 XX  
 AC 21-FEB-2002 (first entry)  
 DT  
 XX  
 XX  
 DE MOL6a protein sequence.  
 KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;  
 KW cell signal processing; metabolic disorder; diabetes; cancer;  
 KW neurodegenerative disorder; immune disorder; cardiac disorder;  
 KW lung disease; autoimmune disease; developmental disorder; antidiabetic;  
 KW Cytostatic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;  
 KW Gene therapy; Vaccine; antiinflammatory; MOL6a; trypsin-like.  
 XX  
 XX Unidentified.  
 OS  
 XX  
 XX  
 PN MO200181578-A2.  
 XX  
 XX  
 PD 01-NOV-2001.  
 XX  
 XX  
 XX 26-APR-2001; 2001WO-US13578.  
 XX  
 XX  
 XX 26-APR-2000; 2000US-200158P.  
 XX  
 XX 28-APR-2000; 2000US-200613P.  
 XX  
 XX 28-APR-2000; 2000US-200780P.  
 XX  
 XX 01-MAY-2000; 2000US-201006P.  
 XX  
 XX 01-MAY-2000; 2000US-201007P.  
 XX  
 XX 01-MAY-2000; 2000US-201236P.  
 XX  
 XX 01-MAY-2000; 2000US-201238P.  
 XX  
 XX 02-MAY-2000; 2000US-201186P.  
 XX  
 XX 03-MAY-2000; 2000US-201474P.  
 XX  
 XX 03-MAY-2000; 2000US-201508P.  
 XX  
 XX 25-JUL-2000; 2000US-220591P.  
 XX  
 XX 15-SEP-2000; 2000US-232678P.  
 XX  
 XX 22-JAN-2001; 2001US-263217P.  
 XX  
 XX 30-JAN-2001; 2001US-265160P.  
 XX  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 XX Vermet CAM, Fernandes ER, Gerlach V, Shimkete RA, Malysankar UM,  
 PI Boldog FL, Zernunen BD, Spytek KA, Majumder K, Tchernov VT;  
 PI Padisaru M, Patturajan M, Burgess CE, Gangoli BA, Smithson G,  
 PI Rastelli L, Macdougall JR, Taupier KO, Grose WM, Szekeres BS;  
 PI Alsdorck JP;  
 XX  
 XX WPI; 2002-049278/06.  
 DR N-PSDB; ABA04594.  
 XX  
 PT Novel G-protein coupled receptor-related polypeptides and  
 PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,  
 PT atherosclerosis, disorders related to cell signal processing and for  
 PT identifying modulators -  
 XX

PS Claim 1; Page 44; 227p; English.

XX The present invention relates to novel G-coupled protein-receptor related  
 CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,  
 CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding  
 CC sequences are useful for treating or preventing a MOLX-associated  
 CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to  
 CC cell signal processing and metabolic pathway modulation, diabetes and  
 CC cancer. Additionally, MOLX proteins and coding sequences are useful for  
 CC preventing and treating a variety of disorders including metabolic  
 CC disorders, nutritional oedema, chronic and hereditary pancreatitis,  
 CC obesity, infectious disease, anorexia, neurodegenerative disorders,  
 CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,  
 CC haematopoietic disorders and various dyslipidemias, metabolic syndrome X  
 CC and wasting disorders associated with chronic diseases and cancers,  
 CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiodysplasia  
 CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,  
 CC multiple sclerosis, lung diseases including asthma, Crohn's disease,  
 CC scleroderma, autoimmune diseases, developmental disorders and neural tube  
 CC defects. The present sequence is the protein sequence for MOL6a.  
 CC MOL6a is a trypsin-like protein.

XX Sequence 235 AA;

Alignment Scores:  
 Pred. No.: 1.53e-126 Length: 235  
 Score: 1255.00 Matches: 235  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 80.45% Indels: 0  
 DB: 23 Gaps: 0

US-10-037-270-482 (1-866) x AAM47664 (1-235)

QY 42 ATGAATATGCTCTTATTTGGGTCCTGCTGGAGCAATTTTCTTCTGACTATCT 101  
 DB 1 MetLysrYrValPheYrLeuGlyValLeuAlaGlyThrPhePheAlaAspSer 20  
 QY 102 GTTCAGAAAGAGACCTGCTCCCTATTTGGTGAACCTCAAGTCTCAACCCCTGT 161  
 DB 21 ValGlnLysGlnuSprProLapProYrLeuValYrLeuLysSerHisPheAsnProCys 40  
 QY 162 GTGGGGCTCTCTCATCAACCCAGCTGGGTCTGGCCCCAGCTCATGCTATTTACCAAT 221  
 DB 41 ValGlyValLeuLeuLysProSerTrpValLeuAlaProAlaHisCysYrLeuProAsn 60  
 QY 222 CTGAAGTGAAGTCTGGGAAATTTCAAGACAGAGTCAAGACGGTATGAACAGCAAT 281  
 DB 61 LeuLysValMetLeuGlnLysPheLysSerArgValArgAspGlyThrGlnInThrIle 80  
 QY 282 AACCCCTTCAGATCGTCGCTACTGTGAATCAATGATGCGCCAGAGATGACCTC 341  
 DB 81 AsnProLleGlnLleValArgYrTrpAsnTrpSerHisSerAlaProGlnAspLeu 100  
 QY 342 ATGCTCATCAAGCTGGCTAAAGCTGCTCATTCATCCAAAGTCCAGCCCTTCCCTC 401  
 DB 101 MetLeuLeuLysLeuAlaLysProLamMetLeuAsnProLysValGlnProLeuProLeu 120  
 QY 402 GCCACCCCAATGTCAGGCGCAAGCACTGTCTGTATCTTCAGGTTTGAAGTGAAGCAA 461  
 DB 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140  
 QY 462 GAAACAGTGGCGAGACCTGACTGGCGAGAACTGAGAGGCCCGCGAGTGTGAT 521  
 DB 141 GluAsnSerGlyArgHisProAspLeuArgGlnAsnLeuValAlaProValMetSerAsp 160  
 QY 522 CGAGAAATGCCAAAAACAGAAAGCAAGAAAGCAAGAAATCTTATGTGAAATTT 581  
 DB 161 ArgGlnCysGlnLysThrGlnGlnGlyLysSerHisArgAsnSerLeuCysValLysPhe 180  
 QY 582 GTGAAGATATTCAGCGCAATTTTGGGAGGTGGCGCTGCTCATCTGATCGCAAGAC 641  
 DB 181 ValLysValPheSerArgLlePheGlyGlnValAlaValAlaThrValIleCysLysAsp 200

QY 642 AACCTCAGGAATTCAGATGGGCGCACTTCATGAGAGGAGCGTGGCATCTACCAAT 701  
 DB 201 LysLeuGlnGlyLleGluValAlaGlyHisPheMetCysGlyLysPheValGlyIleYrThrAsn 220  
 QY 702 GTTTACAAATATGATTCCTGGATTGAGAACACCTGCTTAAGACAG 746  
 DB 221 ValTrpLysrYrValSerTrpLleGlnuAsnThrAlaLysAspLys 235

#### RESULT 4

AAM47665  
 ID AAM47665 standard; Protein; 235 AA.

XX AAM47665;

DT 21-FEB-2002 (first entry)

DE MOL6b protein sequence.

XX MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;  
 KM cell signal processing; metabolic disorder; diabetes; cancer;  
 KM neurodegenerative disorder; immune disorder; cardiac disorder;  
 KM lung disease; autoimmune disease; developmental disorder; antidiabetic;  
 KM Cytostatic; Neutroprotective; Antiatherosclerotic; Immunosuppressive;  
 KM Gene therapy; Vaccine; antiinflammatory; MOL6b; trypsin-like.

XX Unidentified.

PN W0200181578-A2.

XX 01-NOV-2001.

PF 26-APR-2001; 2001MO-US13578.

XX 26-APR-2000; 2000US-200158P.

PR 28-APR-2000; 2000US-200613P.

PR 28-APR-2000; 2000US-200780P.

PR 01-MAY-2000; 2000US-201006P.

PR 01-MAY-2000; 2000US-201007P.

PR 01-MAY-2000; 2000US-201236P.

PR 01-MAY-2000; 2000US-201238P.

PR 02-MAY-2000; 2000US-201186P.

PR 03-MAY-2000; 2000US-201474P.

PR 03-MAY-2000; 2000US-201508P.

PR 25-JUL-2000; 2000US-220591P.

PR 15-SEP-2000; 2000US-232678P.

PR 22-JAN-2001; 2001US-263217P.

PR 30-JAN-2001; 2001US-265160P.

XX (CURA-) CURAGEN CORP.

XX Vermont CAM, Fernandes ER, Gerlach V, Shinkets RA, Maljankar UM;

PI Bolding RU, Zehnuaen BD, Spytek KA, Majumder K, Tcherney VT;

PI Padigaru M, Patlurajan M, Burgess CE, Gangolli BA, Smitheon G;

PI Rastelli L, Macdougall JR, Taupier RJ, Grose WM, Szekeres ES;

PI Alsobrook JP;

XX WPI; 2002-049278/06.

DR N-PSDB; ABA04595.

Novel G-protein coupled receptor-related polypeptides and  
 PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,  
 PT atherosclerosis, disorders related to cell signal processing and for  
 PT identifying modulators -

PS Claim 1; Page 45; 227p; English.

XX The present invention relates to novel G-coupled protein-receptor related  
 CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,  
 CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding  
 CC sequences are useful for treating or preventing a MOLX-associated  
 CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to  
 CC cell signal processing and metabolic pathway modulation, diabetes and





```

QY 162 GTGGGGGCTCTCATCAAAACCCAGCTGGTGTCTGCCCCAGCTCACTGCTATTACCAAAAT 221
DB 41 ValGlyValIleuLeuIleuSerProSerTrpValLeuAlaProIleHisCysTyrLeuProAsn 60
QY 222 CTGAAGAAGTATGCTGGGAAATTTTCAAGACAGAGTCAAGACGGTACTGAACAGCAAAAT 281
DB 61 LeuIleValMetLeuGlyAsnProIleSerArgValArgAspGlyThrGluGlnThrIle 80
QY 282 AACCCCATTCAGATGCTCGCTCACTGGAACCTACAGTCATGAGCGCCCAAGAGTGAACCTC 341
DB 81 AsnProIleGlnIleValArgTyrTrpAsnTyrSerHisSerAlaProGlnAspLeu 100
QY 342 ATGCTCATCAAGCTGCTCAAGCTGCACTGCTCAATCTCAATCCCAAGCTCCCTCCCTC 401
DB 101 MetLeuIleValSerLeuAlaIleProIleAsnProIleValGlnProIleuThrLeu 120
QY 402 GCCACCAACCAATGTCAGGCGGACGCTGTCTGTCTCTCACTGAGTTTGAAGTGAACCAA 461
DB 121 AlaThrThrAsnValArgProGlyThrValCysLeuIleuSerGlyLeuAspTrpSerGln 140
QY 462 GAAACAGTGGCCGACACCTGACTTGGCCGACACCTGGAAGCCCGCCGAGTGTCTGAT 521
DB 141 GluAsnSerGlyArgHisProAspLeuArgGlnAsnLeuGlnAlaProValMetSerAsp 160
QY 522 CGAAGATGCGCAAAAACAGAACAGGAAAAAGCCACAGAAATCTTATGTGTGAATTT 581
DB 161 ArgGluCysGlnIleThrGluGlnGlyLysSerHisArgAsnSerLeuGlyValLysPhe 180
QY 582 GTGAAAGTATTCAGCCGAATTTTGGGAGAGTGGCGCTGCTCACTGCTCACTGCAAAAGAC 641
DB 181 ValIleValIlePheSerArgIlePheGlyGluValAlaValAlaThrValIleCysLysAsp 200
QY 642 AAGCTCCAGGGAATCGAGGTGGGGCACTTCTATGGGAGGGAGACCTCGGCATCTACACCAAT 701
DB 201 LysLeuGlnGlyIleGluValGlyHisPheMetGlyGlyAspValGlyIleTyrThrAsn 220
QY 702 GTTTACAAATATGATCTGATTCGATTGAGAACACACTGCTAAGACCAAG 746
DB 221 ValTyrLysTyrValSerTrpIleGluAsnThrAlaLysAspLys 235

```

```

PT Novel AFP polypeptides and polynucleotides, useful for diagnostic and
PT therapeutic purposes, in cancer therapy and for screening modulator
PT compounds
XX Claim 2, Page 194, 220pp; English.
XX
XX The invention relates to novel human AFP proteins (not defined) and
CC the nucleic acids that encode them. AFP proteins are useful as standards
CC in assays of protein and protein inhibitors in both clinical and research
CC settings, as protein and amino acid supplements, including hydrolyses.
CC The nucleic acids are useful for radiation hybrid mapping. Secretory
CC fusion proteins of AFP are useful in cancer therapy, for enhancing in
CC vitro cytotoxicity, for enhancing in vivo killing of target tissues, and
CC for targeted cells or tissue inhibition or ablation. Anti-AFP antibodies
CC are useful for isolating target polypeptides by affinity purification, in
CC diagnostic assays for determining circulating or localised levels of
CC target polypeptides, for tissue typing, for cell sorting, for screening
CC expression libraries, for generating anti-idiotypic antibodies, and as
CC neutralising antibodies or as antagonists to block protein activity in
CC vitro and in vivo. AFP proteins and nucleic acids may be used to
CC diagnose or treat (e.g. by gene therapy) diseases associated with the
CC malfunction of the AFP e.g. AFP166924 and Marfan's syndrome,
CC AFP576853/AFP39158 and thrombocytopenia, leukaemia, porphyria, Gilles De
CC La Tourette's syndrome, AFP652829 and squamous cell carcinoma, diabetes
CC mellitus, Grave's disease, AFP664311 and colon cancer, AFP251034 and
CC Alzheimer's disease and AFP686580 and epiphyseal dysplasia. Many more
CC examples of disease are given in the specification. The present
CC sequence represents an AFP of the invention.
SQ
XX Sequence 235 AA;
XX
XX Alignment Scores:
Pred. No.: 1,12e-125 Length: 235
Score: 1247.00 Matches: 234
Percent Similarity: 99.57% Conservative: 0
Best Local Similarity: 99.57% Mismatches: 1
Query Match: 79.94% Indels: 0
DB: 22 Gaps: 0
US-10-037-270-482 (1-866) x AAU27666 (1-235)

```











PI Rastelli L, MacDougall JR, Taupier RJ, Grose WM, Szekeres ES;  
 PI Alsbrook JP;  
 XX  
 XX  
 DR WPI: 2002-049278/06.  
 DR N-PSDB; ABA04647.  
 XX  
 PT Novel G-protein coupled receptor-related polypeptides and  
 PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,  
 PT atherosclerosis, disorders related to cell signal processing and for  
 PT identifying modulators -  
 XX  
 XX Example 1; Page 216; 227pp; English.  
 XX  
 CC The present invention relates to novel G-coupled protein-receptor related  
 CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,  
 CC ABA04589-ABA04603 and ABA047659-ABA047673). MOLX proteins and coding  
 CC sequences are useful for treating or preventing a MOLX-associated  
 CC disorder, such as cardiomyopathy, atherosclerosis, disorders related  
 CC to cell signal processing and metabolic pathway modulation, diabetes and  
 CC cancer. Additionally, MOLX proteins and coding sequences are useful for  
 CC preventing and treating a variety of disorders including metabolic  
 CC disorders, nutritional oedema, chronic and hereditary pancreatitis,  
 CC obesity, infectious disease, anorexia, neurodegenerative disorders,  
 CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,  
 CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X  
 CC and wasting disorders associated with chronic diseases and cancers,  
 CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiogenesis  
 CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,  
 CC multiple sclerosis, lung diseases including asthma, Crohn's disease,  
 CC scleroderma, autoimmune diseases, developmental disorders and neural tube  
 CC defects. The present sequence is the protein sequence for a variant of  
 CC MOLA. MOLA is a trypsin-like protein.  
 CC  
 XX  
 SO Sequence 228 AA:  
 Alignment Scores:  
 Pred. No.: 2,46e-122 Length: 228  
 Score: 1216.00 Matches: 226  
 Percent Similarity: 99.12% Conservative: 0  
 Best Local Similarity: 99.12% Mismatches: 2  
 Query Match: 77.95% Indels: 1  
 DB: 23 Gaps: 0  
 US-10-037-270-482 (1-866) x AAM47674 (1-228)  
 QY 42 ATGAATATGCTCTTATTTGGGTGCTGCTGGACATTTTCTTGTGCTGATCTATCT 101  
 DB 1 MetLysTyrValPheTyrLeuGlyValLeuAlaGlyThrPhePheAlaSerSer 20  
 QY 102 GTTCGAAGAAGAGACCTGCTCCCTATTGTTGGTGTACTCAAGCTCACTTCAACCCCTGT 161  
 DB 21 ValGlnLysGlnLysProAlaProTyrLeuValTyrLeuLysSerHisPheAlaProCys 40  
 QY 162 GTGGGGCTCTCATCAACCCAGCTGGGTGCTGGCCCAAGCTCACTGCTATTACCAAT 221  
 DB 41 ValGlyValLeuLeuLeuLysProSerTrpValLeuAlaProAlaHisCysTyrLeuProAla 60  
 QY 222 CTGAAGTATGCTGGGAATTTCAAGAGACAGTCAAGACGCTACTGAAACGACCAAT 281  
 DB 61 LeuLysValMetLeuGlnLysPheLysSerArgValArgPheGlyThrGlnGlnThrIle 80  
 QY 282 AACCCCATTCAGATGCTGGCTAGCTAGCACTAGTACGAGCCCAAGAGATGACCTC 341  
 DB 81 AsnProIleGlnIleValArgTyrTrpAsnTyrSerHisSerAlaProGlnAspLeu 100  
 QY 342 ATGCTCATCAAGCTGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
 DB 101 MetLeuIleLysLeuAlaLysProAlaMetLeuAsnProLysValGlnProLeuThrLeu 120  
 QY 402 GCCACCAACCAATGTCAGGCGGCACTGTGTCTACTCTCAGTTTGAACGAGCCAA 461  
 DB 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140

QY 462 GAAAACAGTGGCCGACACCTTGACTTGGCGGACGAACCTTGAGGCCCTGATGTCTGAT 521  
 DB 141 GlnAsnSerGlyArgHisProAspLeuArgGlnAsnLeuGlnAlaProValMetSerAsp 160  
 QY 522 CGAGAAATGCCA-AAAAACGAAACAAGAAAAAGCCACAGAAATTCCTTATGTGTGAATT 580  
 DB 161 ArgGlnCysGlnLysAsnArgThrArgLysLysProGlnGlnLysPheLeuMetCysGlnIle 180  
 QY 581 TGTGAAGATATTCAAGCCGAATTTTGGGAGGTGGCCGCTGCTACTGTCATCGCAAGA 640  
 DB 181 CysGlnSerIleGlnProAsnPheTrpGlyGlyArgCysTyrCysHisLeuGlnArg 200  
 QY 641 CAAGCTCCAGGGAATCGAGTGGGCACTTCATGGAGGGGAGCTCGGATCTACACCA 700  
 DB 201 GlnAlaProGlyAsnArgGlyGlyValAlaLeuHisGlyArgGlyArgArgHisLeuHisGln 220  
 QY 701 TGTTTTCAAAATATGTATCTTGAT 724  
 DB 221 CysLeuGlnIleCysIleLeuAsp 228  
 RESULT 12  
 AAB21321  
 ID AAB21321 standard; Protein, 247 AA.  
 XX  
 AC AAB21321;  
 XX  
 DT 02-FEB-2001 (first entry)  
 XX  
 DE Human trypsinogen.  
 XX  
 KM Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; trypsinogen;  
 KM kallikrein-like protein; serine protease;  
 KM cytostatic; cancer; prostate cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200053776-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 09-MAR-2000; 2000MO-CA00258.  
 XX  
 PR 11-MAR-1999; 99US-0124260.  
 PR 01-APR-1999; 99US-0127386.  
 PR 21-UTL-1999; 99US-0144919.  
 XX  
 PA (MOUNT ) MOUNT SINAI HOSPITAL.  
 XX  
 PI Yousef GM, Diamandis EP;  
 XX  
 DR WPI, 2000-587440/55.  
 XX  
 XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L  
 PT protein mediated disorders, especially cancer. -  
 PS Example 4; Fig 17; 184pp; English.  
 XX  
 CC The present sequence is human trypsinogen, a member of the serine  
 CC protease family. Kallikreins and kallikrein-like proteins are a  
 CC subgroup of the serine protease enzyme family. They catalyse the  
 CC selective cleavage of specific polypeptide precursors to release  
 CC peptides with potent biological activity. Nucleic acids encoding  
 CC kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5 and  
 CC KLK-L6 have been isolated. The proteins are useful in the treatment,  
 CC monitoring and diagnosis of cancers, especially prostate cancer. They  
 CC can also be used to identify a substance that can associate with or  
 CC mediate the biological activity of the proteins. Antibodies can  
 CC be used to treat conditions mediated by the kallikrein-like proteins.  
 XX  
 SO Sequence 247 AA;  
 Alignment Scores:  
 Pred. No.: 4.93e-29 Length: 247



Db 141 IleserGlyTrrpGlyAsnThrGlnSerIleGlyGlnAsnTyrProAspValLeuGlnCys 160  
 QY 498 CTGAGAGCCCGCGTGATGTGTGATCGAAGATGCCAAAAACAGAACAGGAAAAACCCAC 557  
 Db 161 LeuValAlaProIleLeuSerAspSerValCysArgAsnAlaTyrProGlyGlnIleSer 180  
 QY 558 AGGAATTCCTTATGTGTGAAATTTGTGAAATTTATGACCGAATTTTGGGAGGTGGC 617  
 Db 181 SerAsnMetMetCysLeuGlyTyrMetGlnGlyGlyLeuAspSerCysGlnGlyAspSer 200  
 QY 618 GTTGCTACTGTGATCTGCAAGAACAGCTCCAGGGAATC-----GAGGTGGGAC 668  
 Db 201 GlyGlyProValValCysAsnGlyGlnLeuGlnGlyValValSerTrrpGlyAlaGlyCys 220  
 QY 669 TTCATGGAGGAGGACGTGGCATCTACACCAATGTTTACAAATATGATTCCTGATTGAG 728  
 Db 221 AlaGlnGlyGlyLeuProGlyValSerProGlyValCysLeuTyrValSerTrrpIleGln 240  
 QY 729 AACACT 734  
 Db 241 GlnThr 242  
 RESULT 14  
 AAY31160  
 ID AAY31160 standard; protein; 224 AA.  
 XX AAY31160;  
 AC  
 XX 26-OCT-1999 (first entry)  
 DT  
 XX Human trypsin serine protease protein domain.  
 DE  
 XX  
 KM Macrophage stimulating protein; MSP; human; modulator; proliferation;  
 KW differentiation; intestinal epithelium; colon crypt; treatment; cancer;  
 KW haematopoietic disorder; megakaryocyte deficiency; gastrointestinal;  
 KW chemotherapeutic agent; gut toxicity; serine protease; trypsin.  
 XX  
 OS Homo sapiens.  
 OS  
 PN US594892-A.  
 PD 07-SEP-1999.  
 PF 16-DEC-1996; 96US-0766982.  
 PR 16-DEC-1996; 96US-0766982.  
 PA (AMGB-) AMGEN INC.  
 XX  
 PI Wahl RC;  
 XX  
 DR WPI, 1999-517975/43.  
 XX  
 PT Analogues of macrophage stimulating protein for treating  
 XX gastrointestinal or haematopoietic disorders  
 PS Example 2, Column 27-30, 23pp; English.  
 XX  
 CC This invention describes a novel purified and isolated analogue of mature  
 CC macrophage stimulating protein (MSP) having at least one unpaired  
 CC cysteine residue substituted with another amino acid which modulates the  
 CC proliferation or differentiation of the intestinal epithelium. The  
 CC product of the invention binds to RON (a cell membrane protein tyrosine  
 CC kinase which is a member of the c-met family) to promote the formation of  
 CC colon crypts. MSP analogues are useful for the treatment of conditions  
 CC requiring the administration of MSP, such conditions include  
 CC haematopoietic disorders such as those involving a deficiency of  
 CC megakaryocytes and gastrointestinal disorders such as ulcerative colitis,  
 CC Crohn's disease and infections. The MSP analogues are useful for  
 CC maintaining and repairing the epithelial lining in the treatment of  
 CC cancer, where the aggressive use of chemotherapeutic agents or the use of  
 CC whole body radiation may lead to gut toxicity. The MSP analogues, which  
 CC have a higher activity than normal human MSP are effective at smaller

CC dosages, or optionally, they may be administered less frequently than  
 CC human MSP. This sequence represents a human trypsin serine protease  
 CC domain which is used in a description of the method of the invention.  
 XX  
 SQ Sequence 224 AA;  
 Alignment Scores:  
 Pred. No.: 1e-28 Length: 224  
 Score: 349.50 Matches: 76  
 Percent Similarity: 52.83% Conservative: 36  
 Best Local Similarity: 35.85% Mismatches: 97  
 Query Match: 22.40% Indels: 3  
 DB: 20 Gaps: 1  
 US-10-037-270-482 (1-866) x AAY31160 (1-224)  
 QY 108 AAGAGAACCTGCTCTCTATTTGGTGAATCTCAAGTCTCAACCTTCAACCCCTGTGGC 167  
 Db 8 GlnGlnAsnSerValProTyrGlnValSerLeuAsnSerGlyTyrHisPheCysGlyGly 27  
 QY 168 GTCTCATCAAAACCCAGCTGGGTGGTGGCCGCCAGCTCACTGATTTACCAATCGAAA 227  
 Db 28 SerLeuIleAsnGlnGlnIrrpValValSerAlaGlyHisCysTyrIleSerArgIleGln 47  
 QY 228 GTGATGCTGGAAATTTTCAAGACAGAGTCAAGACGCTACTGAACAGACAAATTAACCC 287  
 Db 48 ValArgLeuGlyGlnHisAsnIleGlnValLeuGlnGlyAsnGlnIleHisAla 67  
 QY 288 ATTGAGATGCTCCGCTACTGTAAGTCAAGTCAATAGCGCCCAAGAGTACCTCATGCTC 347  
 Db 68 AlaValIleIleArgHisProGlnTyrAspArgIleThrLeuAsnAsnAspIleMetLeu 87  
 QY 348 ATCAAGCTGGCTAAGCTCCATCTCAATCCCAAGTCCAGCCCTTCCCTGGCCACC 407  
 Db 88 IleValLeuSerSerArgAlaValIleAsnAlaArgValSerThrIleSerLeuProThr 107  
 QY 408 ACCAATGTGAGCCAGACACTGTCTGTCTACTCTCAGTTTGGACTGAGCCAAAGAAC 467  
 Db 108 AlaProProAlaThrGlyThrIleCysLeuIleSerGlyTrrpGlyAsnThrAlaSerSer 127  
 QY 468 AGTGCCGACACCTTGACTTGGGCGAGAACCTGAGAGCCCGGTGATGTGTGATCGAGA 527  
 Db 128 GlyAlaAspTyrProGlyPheLeuGlnCysLeuAspAlaProValLeuSerGlnAlaLys 147  
 QY 528 TGGCAAAAAACAGAACAGAAAAACCCACAGAAATCTTATGTGTGAAATTTGAAA 587  
 Db 148 CysGlnAlaSerTyrProGlyIleThrIleSerAsnMetPheCysValAlaPheLeuGln 167  
 QY 588 GTATTCAGCCGAATTTTGGGGAGGTGGCGTGTGCTACTGTCATCTGCAAAAGACAGCTC 647  
 Db 168 GlyGlyIleAspSerCysGlnGlyAspSerGlyGlyProValValCysAsnGlnLeu 187  
 QY 648 CAGGAAATCAGAGTG-----GGGCACTTCATGAGGAGGAGCGCTCGCATCTACCC 698  
 Db 188 GlnGlyValValSerTrrpGlyAspGlyCysAlaGlnIleAsnIleAsnIleProGlyValTyrThr 207  
 QY 699 AATGTTTACAAATATGTATCTGATTCGATTGGAACACT 734  
 Db 208 LysValTyrAsnTyrValIleTrrpIleLysAsnThr 219  
 RESULT 15  
 AAB98503  
 ID AAB98503 standard; Protein; 225 AA.  
 XX AAB98503;  
 AC  
 XX 03-AUG-2001 (first entry)  
 DT  
 XX Human trypsin serine protease catalytic domain.  
 DE  
 XX Human, TADG-15, cytostatic; vaccine; ovarian tumour; cancer; trypsin;  
 KW tumour antigen-derived gene 15; serine protease.

OS Homo sapiens.  
 XX MO200129056-A1.  
 XX 26-APR-2001.  
 XX 20-OCT-2000; 2000MO-US29095.  
 XX 20-OCT-1999; 99US-0421213.  
 XX (UYAR-) UNIV ARKANSAS.  
 XX O'Brien TJ, Tanimoto H;  
 XX WPI; 2001-381031/40.  
 XX  
 XX Novel extracellular serine protease, termed tumor antigen-derived gene  
 PT 15 protein overexpressed in carcinomas and DNA encoding it, for  
 PT diagnosis, treatment, prevention of cancer, particularly breast,  
 PT ovarian cancer -  
 XX  
 XX Example 10; Fig 1; 130pp; English.  
 XX  
 XX The present invention relates to human tumor antigen-derived gene 15  
 CC (TADG-15) protein and coding sequence (see AAB23601 and AAB98500).  
 CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is  
 CC over-expressed in ovarian tumors. TADG-15 protein or its fragments of  
 CC 9-20 residues that lack TADG-15 protease activity are useful for  
 CC vaccinating an individual against TADG-15, having, suspected of having or  
 CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a  
 CC diagnostic or therapeutic target in cancer. The present sequence was used  
 CC in a sequence homology alignment with the catalytic domain of TADG-15.  
 XX  
 XX Sequence 225 AA;

Alignment Scores:  
 Pred. No.: 1e-28 Length: 225  
 Score: 349.50 Matches: 76  
 Percent Similarity: 52.83% Conservative: 36  
 Best Local Similarity: 35.85% Mismatches: 97  
 Query Match: 22.40% Indels: 3  
 DB: 22 Gaps: 1

US-10-037-270-482 (1-866) x AAB98503 (1-225)

QY 108 AAGAGACCCCTGCTGCTATTTGGTACCTCAAGTCTCACTCAACCCCTGTGGGC 167  
 Db 9 GUGLUNASERVALPROTYRGINVALSERLEUANSERGLYTRHISNECYSGLYGLY 28  
 QY 168 GTCCCTCATCAACCCAGCTGGGTGCTGGCCCAAGCTCACTGCTATTACCAATCTGAAA 227  
 Db 29 SERLEULEANSGLUGINTRPVALISERVALGLYHISCYSTYRLEYSERARGILEGLN 48  
 QY 228 GTGATGCTGGGAATTTCAGAGCAGAGTCAAGACCGTACTGACGACGACATTAACCC 287  
 Db 49 VALARGLEULIYGLUHISANILIEGLUVALLEUNGILYASNGILNPHIEASNALA 68  
 QY 288 ATTCAGATCTCCGCTACTGGAAGTCAAGTCAATGAGCCCAAGTCAAGTCAAGTCAAGTCA 347  
 Db 69 ALALYSELLELLEIGHISPROGINTRYASPARGLYTHRLEUANSANASPRILEMEULEU 88  
 QY 348 ATCAAGCTGCTAAGCTGCTGCTCAATCCCAAGTCAAGCCCTTCCCTGCGCACC 407  
 Db 89 ILELYSELEUSERARGALAVALLIEASNALARGVALSERTHRILESERLEUPROTHR 108  
 QY 408 ACCAATGTACGGCAGGCACTGTCTCTACTCTCAAGTTTGGACGTGAGCCAAAGAAAC 467  
 Db 109 ALAPROPRALATHRGILYTHIRLYSCYLEULILESERGLYTRPGLYANTRHALASER 128  
 QY 468 AGTGCCGACACCTGACTTGCGCGAAGTCAAGTCAAGCCCGGAGTGTCTGATCGAGAA 527  
 Db 129 GLYALASPTYRPROASPGILULEUGINCYLEUASPALAPROVALLEUSERGINALALYS 148

QY 528 TGCCAAAAACAGAACAGAAAAAGCCACAGAAATTCCTTATGTGTGAATTTGTGAAA 587  
 Db 149 CYEGULALASERTYRPROGLYLYSILETHRSERASMETPHECYVALIGLYPHELEUGLU 168  
 QY 588 GTATTACGCCGAATTTTGGGAGGTGGCCGTGCTACTCTCATCTGCCAAGACAGCTC 647  
 Db 169 GLYGLYLYSAPSERCYSGINGLYASPSERGLYGLYPROVALVALCYASANGLYGLNLEU 188  
 QY 648 CAGGAAATCGAGGNG-----GGGCACTTCATGGAGGAGGACGTGGCATCTACACC 698  
 Db 189 GINGLYVALVALSERTRPGIYASPGIYCYVALAGILYASANLYSPROGLYVALTYRTHR 208  
 QY 699 AATGTTTACAATATATGATCTGATCTGATTTGAGAACT 734  
 Db 209 LYSVALTYRASNTRYRVALLYSTRPILEYASANTHR 220

Search completed: November 8, 2003, 02:08:08  
 Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: November 8, 2003, 02:06:22 / Search time 19 Seconds  
(without alignments)  
3856.969 Million cell updates/sec

Title: US-10-037-270-482  
Perfect score: 1560  
Sequence: 1 ggcacttaccctgcagccta.....tgggaatgcagcactagct 866

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 328717 seqs, 42310858 residues  
Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DRV=xlh  
-Q=/cgn2\_1/USFTO.spool/US10037270/runat.07112003.140516.8822/app\_query.fasta\_1.1031  
-DB=Issued\_Patents\_AA -QFMT=fasta -SUFFIX=rat -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10037270 @CNC1.1.18 @runat.07112003.140516.8822 -NCPU=6 -ICPU=3  
-NO\_MMAP -LANG=QUERY -NEG\_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/POCUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1247	79.9	235	4	US-09-796-110-2
2	349.5	22.4	224	2	US-08-766-982-13
3	349.5	22.4	224	2	US-08-944-483-36
4	349.5	22.4	224	3	US-09-296-219-13
5	349.5	22.4	225	2	US-09-027-337-5
6	349.5	22.4	225	4	US-09-644-600-5
7	346.5	22.2	246	2	US-08-978-404B-44
8	344.5	22.1	224	3	US-08-944-483-35
9	335.5	21.5	247	2	US-08-956-267A-2
10	334.5	21.4	224	2	US-08-944-483-34
11	319.5	20.5	281	1	US-08-467-155A-7
12	319.5	20.5	281	2	US-08-628-198-7

13	319.5	20.5	281	3	US-09-201-038-7	Sequence 7, Appl1
14	319.5	20.5	281	5	PCT-US96-07343-7	Sequence 7, Appl1
15	312.5	20.0	223	1	US-08-278-091-9	Sequence 9, Appl1
16	312.5	20.0	223	1	US-08-483-859-9	Sequence 9, Appl1
17	312.5	20.0	223	1	US-08-472-173-9	Sequence 9, Appl1
18	312.5	20.0	223	2	US-08-487-167-9	Sequence 9, Appl1
19	312.5	20.0	223	2	US-08-482-816-9	Sequence 9, Appl1
20	312.5	20.0	223	2	US-08-296-149-9	Sequence 9, Appl1
21	312.5	20.0	223	2	US-08-801-499-9	Sequence 9, Appl1
22	312.5	20.0	223	2	US-08-615-271-9	Sequence 9, Appl1
23	312.5	20.0	223	3	US-09-074-660-9	Sequence 9, Appl1
24	312.5	20.0	223	3	US-09-074-659-9	Sequence 9, Appl1
25	312.5	20.0	223	3	US-09-106-468-9	Sequence 9, Appl1
26	312.5	20.0	223	3	US-09-106-466A-9	Sequence 9, Appl1
27	312.5	20.0	223	3	US-09-106-467-9	Sequence 9, Appl1
28	312.5	20.0	229	3	US-09-120-582-2	Sequence 2, Appl1
29	304.5	19.5	299	1	US-08-467-155A-8	Sequence 8, Appl1
30	304.5	19.5	299	2	US-08-628-198-8	Sequence 8, Appl1
31	304.5	19.5	299	3	US-09-201-038-8	Sequence 8, Appl1
32	304.5	19.5	299	5	PCT-US96-07343-8	Sequence 8, Appl1
33	287.5	18.4	244	1	US-08-361-395-1	Sequence 1, Appl1
34	280	17.9	286	1	US-08-467-155A-9	Sequence 9, Appl1
35	280	17.9	286	2	US-08-628-198-9	Sequence 9, Appl1
36	280	17.9	286	3	US-09-201-038-9	Sequence 9, Appl1
37	280	17.9	286	5	PCT-US96-07343-9	Sequence 9, Appl1
38	261	16.7	156	3	US-09-261-416-6	Sequence 6, Appl1
39	252	16.2	260	3	US-09-008-271A-7	Sequence 7, Appl1
40	251	16.1	260	3	US-09-070-526-2	Sequence 2, Appl1
41	250	16.0	288	4	US-09-386-642-13	Sequence 13, Appl1
42	249	16.0	271	1	US-08-467-155A-10	Sequence 10, Appl1
43	249	16.0	271	2	US-08-628-198-10	Sequence 10, Appl1
44	249	16.0	271	3	US-09-201-038-10	Sequence 10, Appl1
45	249	16.0	271	5	PCT-US96-07343-10	Sequence 10, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-796-110-2  
; Sequence 2, Application US/09796110  
; Patent No. 6468776  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell C.  
; TITLE OF INVENTION: Human Serine Protease  
; FILE REFERENCE: 00-16  
; CURRENT APPLICATION NUMBER: US/09/796,110  
; CURRENT FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-796-110-2

#### Alignment Scores:

Pred. No.: 4,236-132  
Score: 1247.00  
Percent Similarity: 99.57%  
Best Local Similarity: 99.57%  
Query Match: 79.94%  
DB: 4  
Length: 235  
Matches: 234  
Conservative: 0  
Mismatches: 1  
Indels: 0  
Gaps: 0

US-10-037-270-482 (1-866) x US-09-796-110-2 (1-235)

QY 42 ATGAATATGCTCTTCATTTGGTGGCTCGCGGACATTTTCTTGTGACTCATCT 101  
DB 1 MchLrYrValPhetYrIeuSlYalIeuAlaGlymPhenPheNlaAspSer 20  
QY 102 GTTCAGAAAGAACCCCTGCTCCCTATTGTTGTTGACCTCAAGCTTCACCTCAACCCCTGT 161  
DB 21 ValGlnYeuGluAspProIlaProYrIeuValYrIeuYSerIlaPheAsnProCyS 40

```

QY      162 GTGGCGCTCCATCAAAACCAGCTGGGTGGCCGCCAGCTACCTATTATTAACCAAT 221
      41 ValGlyValLeuIleLysProSerTrpValLeuAlaProIleHisCysTrpLeuProAsn 60
QY      222 CTGAAAGTAGTGTGGGAAATTTCAAGACAGACAGACTGACGAGTCTGAAACAGCAATT 281
      61 LeuLysValMetLeuGlyAsnPheLysSerArgValArgAspGlyThrGluGlnThrIle 80
QY      282 AACCCCATTCAGATCGCCGCTACGTGACATCAAGTATGCGGCCCCACAGAGTAGACTC 341
      81 AsnProIleGlnIleValArgTrpAsnTrpSerHisSerAlaProGlnAspAspLeu 100
QY      342 ATGCTCATCAAGCTGGCTAAGCCTGACCTGACCTCAATCCCAAGTCCAGCCCTTCCCTC 401
      101 MetLeuIleLysLeuAlaLysProIleMetLeuAsnProLysValGlnProLeuThrLeu 120
QY      402 GCCACCACTATGTCAAGCCAGGACACTGTGTCTACTCTCAAGTTTGAAGTGAACCAA 461
      121 AlaThrThrAsnValArgProGlyThrValLysLeuLeuSerGlyLeuAspTrpSerGln 140
QY      462 GAAACAGTGGCCGACACCTGACCTTGGCGAGAACCTGAGAGCCCGCGATGCTGAT 521
      141 GluAsnSerLysArgHisProAspLeuArgGlnAsnLeuGlnAlaProValMetSerAsp 160
QY      522 CGAATATGCCAAAACAGACAAAGAAAACCCACAGAAATTCCTTATGTGAAATTT 581
      161 ArgGluCysGlnLysThrGluGlnGlyLysSerHisArgAsnSerLeuCysValLysPhe 180
QY      582 GTGAAATGATTACGCCGAATTTTGGGAGGTGGCCGTTGCTACTGATCTGCAAAAGAC 641
      181 ValLysValIlePheSerArgLysIlePheGlyGluValAlaValAlaThrValIleCysValAsp 200
QY      642 AACCTCAGAGGAATCGAGGTGGGCACTTCACTGAGAGGAGGACGTCGCAATCAACCAAT 701
      201 LysLeuGlnIleLysIleGluValGlyHisPheMetCysLysValArgValGlyIleTrpAsn 220
QY      702 GTTACAAATATGATCTGCTGATGAGAACACTGCTAAGACAAG 746
      221 ValTrpLysTrpValSerTrpIleGlnAsnThrAlaLysAspLys 235
      Db
      RESULT 2
      US-08-766-982-13
      ; Sequence 13, Application US/08766982
      ; Patent No. 5948892
      ; GENERAL INFORMATION:
      ; APPLICANT: Wahl, Robert C.
      ; TITLE OF INVENTION: Analogs of Macrophage Stimulating
      ; NUMBER OF SEQUENCES: 13
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSEE: Amgen Inc.
      ; STREET: 1840 Denavilland Drive
      ; CITY: Thousand Oaks
      ; STATE: California
      ; COUNTRY: USA
      ; ZIP: 91320-1789
      ; MEDIUM TYPE: Floppy disk
      ; COMPUTER: IBM PC compatible
      ; OPERATING SYSTEM: PC-DOS/MS-DOS
      ; SOFTWARE: PatentIn Release #1.0, Version #1.30
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: US/08/766,982
      ; FILING DATE:
      ; CLASSIFICATION: 536
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: Winter, Robert B.
      ; REFERENCE/DOCKET NUMBER: A-441
      ; INFORMATION FOR SEQ ID NO: 13:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 224 amino acids
      ; TYPE: amino acid
      ; STRANDEDNESS: single

```

```

      ; TOPOLOGY: linear
      ; MOLECULE TYPE: protein
      ; US-08-766-982-13
      Alignment Scores:
      Pred. No.: 8,44e-31 Length: 224
      Score: 349.50 Matches: 76
      Percent Similarity: 52.83% Conservative: 36
      Best Local Similarity: 35.85% Mismatches: 97
      Query Match: 22,40% Indels: 3
      Gaps: 1
      US-10-037-270-482 (1-866) x US-08-766-982-13 (1-224)
      QY      108 AAGAAGACCCCTGCTCTATTGTTGTTGTAACCTCAAGTCTCACTTCAACCCCTGTGGG 167
      8 GluGlnAsnSerValProTyrGlnValSerLeuAsnSerGlyTrpHisPheCysGlyGly 27
      Db      168 GTCTCATCAAAACCCAGCTGGGTGGCCGCCAGCTCACTATTATTAACCAATCTGAA 227
      28 SerLeuIleAsnGluGlnTrpValValSerAlaGlyHisCysTrpLysSerArgIleGln 47
      QY      228 GTGATGCTGGGAAATTTCAAGACAGATCAGACGCTACTGACAGACAAATTAAACCC 287
      48 ValArgLeuGlyGluHisAsnIleGluValLeuGluGlyAsnGluGlnPheIleAsnAla 67
      Db      288 ATTCAAGATGCTCGCTACTGAACTACAGTCATAGGCCCCACAGAGATGACCTCATGCTC 347
      68 AlaLysIleLeuThrArgHisProGlnTrpAspArgLysThrLeuAsnAsnAspIleMetLeu 87
      QY      348 ATCAACTGCTGAAGCTGCTCACTGATCTCAATCCCAAGTCCAGCCCTTCCCTCGCCACC 407
      88 IleLysLeuSerSerArgHisAlaValIleAsnAlaArgValSerThrIleSerLeuProThr 107
      Db      408 ACCAATGTACAGCCAGGCACTGTGTCTACTCTCACTGAGTTTGAAGTGAACCAAGAAAC 467
      108 AlaProProAlaThrGlyThrLysCysLeuIleSerGlyTrpLysAsnThrAlaSerSer 127
      QY      468 AGTGGCGACACCTGACTGCTGGCGAGAACCTGAGAGCCCGCGTGAATGTCATCGAGAA 527
      128 GlyAlaAspTrpProAspGluLeuGlnCysLeuAspAlaProValLeuSerGlnAlaLys 147
      Db      528 TGCCAAAACAGACAAAGAAAACCCACAGAAATTCCTTATGTGAAATTTGAA 587
      148 CysGlnAlaSerTrpProGlyLysIleThrSerAsnMetPheCysValGlyPheLeuGln 167
      QY      588 GTATTACGCCGAATTTTGGGAGGTGGCCGTTGCTACTGTCACTGCAAAAGCAAGCTC 647
      168 GlyGlyLysAspSerCysGlnGlyAspSerGlyGlyProValValCysAsnGlyGlnLeu 187
      Db      648 CAGGAATCGAGTG-----GGGCACTTCATGGAGGGGACGTCGCACTTCACAC 698
      188 GlnGlyValValSerTrpGlyAspGlyCysAlaGlnLysAsnLysProGlyValTrpThr 207
      QY      699 AATGTTACAAATATGATCTGCTGATGAGAACACT 734
      208 LysValLysAsnTrpValLysTrpIleLysAsnThr 219
      Db
      RESULT 3
      US-08-944-483-36
      ; Sequence 36, Application US/08944483
      ; Patent No. 6232456
      ; GENERAL INFORMATION:
      ; APPLICANT: COHEN, MAURICE
      ; APPLICANT: COLETTIS, TRACEY L.
      ; APPLICANT: FRIEDMAN, PAULIA N.
      ; APPLICANT: GRANADOS, EDWARD N.
      ; APPLICANT: KASS, MICHAEL R.
      ; APPLICANT: RUSSELL, JOHN C.
      ; APPLICANT: STEWART, KENT D.
      ; APPLICANT: STROUPE, STEVEN D.
      ; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
      ; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

```

TITLE OF INVENTION: OF THE PROSTATE  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 62324566  
US-08-944-483-36  
Alignment Scores:  
Pred. No.: 8.44e-31 Length: 224  
Score: 349.50 Matches: 76  
Percent Similarity: 52.83% Conservative: 36  
Best Local Similarity: 35.85% Mismatches: 97  
Query Match: 22.40% Indels: 3  
DB: 3 Gaps: 1  
US-10-037-270-482 (1-866) x US-08-944-483-36 (1-224)  
QY 108 AAGAAGACCCGCTGCTGCTATTGGTACCTCAAGTCTCACTTCAACCCGTGTGGGC 167  
DB 8 GUGUAASerValProTyrGlnValSerLeuAsnSerGlyTyrHisPheCysGlyGly 27  
QY 168 GTCTTCATCAAAACCCAGCTGGTGTGCGCCCACTCACTGCTATTACCAATCTGAAA 227  
DB 28 SerLeuIleAsnGlnGlnTrpValValSerAlaGlyHisCysEtyrLysSerArgIleGln 47  
QY 228 GTGATCTCGGGAATTTCAAGACAGACAGTCAAGAGCTGACTCAACAATTAACCC 287  
DB 48 ValArgLeuGlyGlyGlnHisAsnIleGlnValLeuGlnGlyAsnGlnGlnInPheIleAsnAla 67  
QY 288 ATTCAAGTGTCCGCGCACTGAGAACTACAGTCATAGAGCCCAAGAGATGACCTCATGCTC 347  
DB 68 AlAlValSerIleLeuArgHisProGlnTyrAspArgLysThrLeuAsnAsnAspIleMetLeu 87  
QY 348 ATCAAGCTGCTAGAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTCGGCACC 407  
DB 88 IleLysLeuSerSerArgAlaValIleAsnAlaArgValSerThrIleSerLeuProThr 107  
QY 408 ACCAATGTACAGCCAGGCACTGTCTGTCTACTCAAGTTTGACGTGAGCCAAAGAAAC 467  
DB 108 AlaProProAlaThrArgLysThrLysCysLeuIleSerGlyTyrpLysAsnThrAlaSerSer 127  
QY 468 AGTGGCGACACACCTGACTTGGCGGAGAACCTGAGGCCCCCGTGAATGCTGATCGAGA 527

DB 128 GlyAlaAspTyrProAspGlnLeuGlnCysLeuAspAlaProValLeuSerGlnAlaLys 147  
QY 528 TGGCAAAAAACAGAACAGAAAAGCCACAGAAATTCCTTGTGTGAAATTTGGAAA 587  
DB 148 CysGlnAlaSerTyrProGlyLysIleThrSerAsnMetPheCysValGlyPheLeuGln 167  
QY 588 GTATTCAAGCCGAATTTTGGGAGGCGCTTGTCTACTGCTCATGTCMAAGAACAGCTC 647  
DB 168 GlyGlyLysAspSerCysGlnGlyAspSerGlyGlyProValValCysAsnGlyGlnLeu 187  
QY 648 CAGGAATCGAGTG-----GGCACTTCATGGAGGAGGAGCTGCGATCTACACC 698  
DB 188 GlnGlyValValSerTyrpGlyAspGlyCysAlaGlnLysAsnLysProGlyValTyrThr 207  
QY 699 AATGTTTACAAATATGATCTCCGATTCGATGAGAACACT 734  
DB 208 LysValTyrAsnTyrValLysTrpIleLysAsnThr 219  
RESULT 4  
US-09-296-219-13  
Sequence 13, Application US/09296219  
Patent No. 6248560  
GENERAL INFORMATION:  
APPLICANT: Wahl, Robert C.  
TITLE OF INVENTION: Analogs of Macrophage Stimulating  
PROTEIN  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavenland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/296,219  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-441  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-296-219-13  
Alignment Scores:  
Pred. No.: 8.44e-31 Length: 224  
Score: 349.50 Matches: 76  
Percent Similarity: 52.83% Conservative: 36  
Best Local Similarity: 35.85% Mismatches: 97  
Query Match: 22.40% Indels: 3  
DB: 3 Gaps: 1  
US-10-037-270-482 (1-866) x US-09-296-219-13 (1-224)  
QY 108 AAGAAGACCCGCTGCTGCTATTGGTACCTCAAGTCTCACTTCAACCCGTGTGGGC 167  
DB 8 GUGUAASerValProTyrGlnValSerLeuAsnSerGlyTyrHisPheCysGlyGly 27  
QY 168 GTCTTCATCAAAACCCAGCTGGTGTGCGCCCACTCACTGCTATTACCAATCTGAAA 227  
DB 28 SerLeuIleAsnGlnGlnTrpValValSerAlaGlyHisCysEtyrLysSerArgIleGln 47



[illegible]

```

RESULT 5
US-09-027-337-5
; Sequence 5, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 5
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Unknown
FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of trypsin (Try)
US-09-027-337-5

Alignment Scores:
Pred. No.:      8,46e-31      Length:      225
Score:          349.50        Matches:      76
Percent Similarity: 52.83%    Conservaive:  36
Best Local Similarity: 35.85%  Mismatches:  97
Query Match:    22.40%       Indels:       3
DB:             2           Gaps:         1

US-10-037-270-482 (1-866) x US-09-027-337-5 (1-225)

Oy      108  AAGAAGACCCCTGCTCCCATTTGGGTGCTACAGCTCACTTCAACCCCTGTGGGCG 167
Db      9   GUGUGAaSeRvAlPrOtyrGlnValSerLeuAaSeRGIYtyrIdSPheCysGIyGLy 28

```

Oy	168	GTCCTCATCAAAACCCAGCCTGGGTGCTGGGCCCGCCAGCTCATCTGATTTACCACCAATCGAAA	227
Dd	29	SerLeuIleAenGIuGlntPrValValSerAlaAglHisCyTylLysSerArgIleGln	48
Oy	228	GTGATGCTGGGAATTTCMAAGCAGAGCTCAGAAGCCGTAATAACAAGCAATTAAACCCC	287
Dd	49	ValArgLeuGIuGlntHieAsnIleGIuValLeuGIuGlnIleAsnGIuInPheIleAenAla	68
Oy	288	ATTGAGATGTCCTCCGTACTGTGGAACCTACAGCTCATACCGCCCCACAGATTAACCTCACAGCTC	347
Dd	69	AlaLysVtIleIAeAghIsrProGIuTrArpArgLyVsThrLeuAsnAsnAspIleWetLeu	88
Oy	348	ATCAAGCTGGTGAGCCCTGCATCGCATCTCAATCCCAAAGTCCAGCCCTTCCTCCCTCCGACCC	407
Dd	89	IleLysLeuSerSerArgAlaValIleAsnAlaArgValSerThrIleSerLeuProThr	108
Oy	408	ACCAATGTCAGGCCAGGACGTGTCTGTACTCTCAGGTTTGAGCTGAGCCAAAGCAAAAAAC	467
Dd	109	AlAProFolAtHngILThrlYcCysLeuIleSerGIuTyrgLIyAsnThrIAsrSer	128
Oy	468	AGTGGCCGACACCTTGACTTGGCGGAGAACCTGGAGAGCCCCCGTAGTGTCTGATCGAGAA	527
Dd	129	GLyAlaAspTYrProAspGIuLeuGlnCYusLeuAspAlaProValLeuSerGlnAlaLys	148
Oy	528	TGCCAAAAACAAGAACAGAAAAAACCACAGAGATTCCTTAATGTGTGAAATTTTGAAA	587
Dd	149	CysGIuAlaSerTYrProGIuLysIleHrsrAsnMerPhcCYuValGIyPheLeuGIu	168
Oy	588	GTATTCAGCCGAATTTTTGGGAGGTGGCCCTGTGTAAGTATCTGCATCTGCMAAGACAAGCTC	647
Dd	169	GIyGIyLysAspSerCYeGIuGIyAspBerGIyGIyProValValCYuAsnGIuGlnLeu	188
Oy	648	CAGGGAATCGAGGTC-----GGGCACTTACATGGGAGGGGAGAGCTGGCATCTACAC	698
Dd	189	GIuGIyAlaValAsrTrpGIyAspLIyCYuAlaGlnLYuAsnLYuAspOLyValTYrThr	208
Oy	699	AATGTTTACAATATGATNCCTGCATTCGAATGACAACACT	734
Dd	209	LysValTYrAsnTYrValLYuValTrpIleLysAsnThr	220

RESULT 6	
US-09-644-600-5	
Sequence 5, Application US/09644600	
Parent No. 6451500	
GENERAL INFORMATION:	
APPLICANT: O'Brien, Timothy J.	
APPLICANT: Tanimoto, Hirotsoshi	
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease	
TITLE OF INVENTION: Overexpressed in Carcinomas	
FILE REFERENCE: D6064CIP/D	
CURRENT APPLICATION NUMBER: US/09/644,600	
CURRENT FILING DATE: 2000-08-23	
PRIOR APPLICATION NUMBER: 09/421,213	
PRIOR FILING DATE: 1999-10-20	
PRIOR APPLICATION NUMBER: 09/027,337	
PRIOR FILING DATE: 1998-02-20	
NUMBER OF SEQ. ID NOS: 98	
SEQ ID NO 5	
LENGTH: 225	
TYPE: PRT	
ORGANISM: Homo sapiens	
FEATURE:	
OTHER INFORMATION: Trypsin	
US-09-644-600-5	
Alignment Scores:	
Pred. No.: 8,466-31	Length: 225
Score: 349.50	Matches: 76
Percent Similarity: 52.83%	Conservative: 36
Best Local Similarity: 35.85%	Mismatches: 97
Query Match: 22,40%	Indels: 3
DB: 4	Gaps: 1



US-10-037-270-482 (1-866) x US-09-644-600-5 (1-225)

QY 108 AAGAGACCCCTGCTCCCTATTGGTGTGTAACCTCAAGTCTCACTTCAACCCCTGTGTGACC 167  
Db 9 GUGUUAASerValProTyrGlnValSerLeuAenSerGlyTyrIleAphCysGly 28  
QY 168 GTCTCATTAACACCCAGCTGGGTGGGCCCCAGCTCACTGTATTATTAACCAATCTGAAA 227  
Db 29 SerLeuIleAenGlnGlnTyrValValSerIleAglTyrIleCysTyrIleSerArgIleGln 48  
QY 228 GTGATGCTGGGAAATTTCAAGACGACAGAGTCACAGACGGGTACTGTAAACAGACAATTAACCCC 287  
Db 49 ValArgLeuGlyGlnHisAenIleGlyValIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 68  
QY 288 ATTACAGATGTCCTCGCTACTGTGAATCACTCACTCACTCACTCACTCACTCACTCACTCACTC 347  
Db 69 AlaIleIleIleIleArgHisIleProGlnTyrIleAparGlyIleThrLeuAenAenAparIleMetLeu 88  
QY 348 ATCAACCTGGCTTAAGCTTGGCTCCATGCTCAATATCCCAAGTCCAGCCCTTCCCTCGCACCC 407  
Db 89 IleIleLeuSerSerArgIleValIleAenIleAenIleAenIleSerThrIleSerLeuProThr 108  
QY 408 ACCAATGTCAAGCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467  
Db 109 AlaProProAlaIleThrGlyIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 128  
QY 468 AGTGGCCGACACCTGTACTGCGGAGAACCTGTGAGGCCCCGTGATGCTGTGATCGAGA 527  
Db 129 GlyAlaAspTyrProArgPglIleuGlnGlnCysLeuAenAparIleProValIleuSerGlnAlaIle 148  
QY 528 TGGCAAAAAACAGACACAGAAAAAACCCACAGAAATTCCTATGTGTGAAATTTGTGAAA 587  
Db 149 CysGlnAlaSerTyrProGlyIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 168  
QY 588 GTATTCAGCCGAATTTTGGGGAGGCGCCGTGTCTACTGTCATCTGATCGAACAAGACAGCTC 647  
Db 169 GlyGlyIleAspSerCysGlnIleAspSerGlyGlyProValIleValCysAenGlyGlnIleu 188  
QY 648 CAGGAATCGAGTG-----GGGCACTTCATGGAGGGAGCGTGGCATCATCACCC 698  
Db 189 GlnGlyValValSerTyrPglIleAspGlyCysAlaGlnIleuAenIleProGlyValIleThr 208  
QY 699 AATGTTTCAATATGTATCTGATTCGATTTGAAACACT 734  
Db 209 LysValTyrAsnTyrValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 220

RESULT 7

US-08-978-404B-44

Sequence 44, Application US/08978404B

Patent No. 5968782

GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.

TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES

TITLE OF INVENTION: FIBRINOGEN

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacke, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/978,404B

FILING DATE: 25-NOV-97

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/032,354

```

FILEIN: DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 5968782e
US-08-978-404B-44

Alignment Scores:
Pred. No.: 1,93e-30 Length: 246
Score: 346.50 Matches: 76
Percent Similarity: 54.50% Conservative: 79
Best local Similarity: 36.02% Mismatches: 93
Query Match: 22.21% Indels: 3
DB: Gaps: 1

US-10-037-270-482 (1-866) x US-08-978-404B-44 (1-246)
QY 111 GAAGACCCCTGCTCCATTTGGTGTACCTCAAGTCTCACTCAACCCCTGCTGGGCGTC 170
DB 32 GlnHisSerValProTyrGlnValSerLeuIleuSerGlyTyrHisPheCysGlyIleSer 51
QY 171 CTCATCAAAACCCAGCTGGGTGCTGGCCCCAGCTCACTCTATTACCAAACTGMAAGTG 236
DB 52 LeuIleAsnAspGlnThrValValSerLalaHisCysGlyLysSerArgIleGlnVal 71
QY 231 ATGCTGGGAAATTTCAAGACGAGTCAAGACGCTACTGAAACGACAATTACCCCAATT 290
DB 72 ArgLeuGlyGlnHisIleAsnIleAsnValLeuGlnGlyAspGlnGlnPheIleAsnLala 91
QY 291 CAATGTCGCGCTACGTAAGTCAACGATCAAGCCGCGCCACAGATGACCTGATGCTATC 350
DB 92 LysIleIleLeuHisIleProAsnTyrSerSerTrpThrLeuAsnAsnAspIleMetLeuIle 119
QY 351 AACCTGCTAAGCTGCTCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGCGCCACACC 410
DB 112 LysLeuSerSerProValIleAsnLalaValLalaProValLalaLeuProSerLala 133
QY 411 AATGTACGCGCAGGCACTGTCTGTCTACTCTAGGTTTGAACCTGAGCCCAAGAAACAGT 470
DB 132 CysAlaProAlaGlyThrGlnCysLeuIleSerGlyTrpGlyAsnThrLeuSerAsnGly 151
QY 471 GCGCGACACCTTGACTTGGGGAGAAACCTCGAAGCGCCCGGTGATGTCGATCGAAGATGC 530
DB 152 ValAsnAsnProAspLeuLeuGlnCysValAlaProValLeuSerGlnAlaAspCys 177
QY 531 CAAAAAACAAGAACAGAAAGAAAGCCACAGCAAGATTCCTATGTCTGAATTTGGAAAGTA 590
DB 172 GluAlaAlaIleTyrProGlyIleIleHisSerSerMetIleCysValIleGlyPheLeuGlnGly 199
QY 591 TTCAGCGCAATTTTGGGGAGGTGCCCTGTCTACTGATCTGATCTGCAAGACAGCTTCAG 650
DB 192 GlyLysAspSerCysGlnGlnLysAspGergGlyGlyProValIleCysAsnGlnGlnLeuGln 211
QY 651 GGAATC-----GAGTGGGCGCATTCATGGAGGGAGCGTCGCGATCTTAAACCAAT 701
DB 212 GlyIleValSerTrpGlyTyrGlyCysAlaLeuProAsnProAsnProGlyValTyrThrLys 231
QY 702 GTTACAAATATGATCTCTGATGGAACCACT 734
DB 232 ValCysAsnProValGlyTyrIleGlnAspThr 242

```

US-08-944-483-35  
Sequence 35, Application US/08944483  
Patent No. 6232456  
GENERAL INFORMATION:  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: KLAS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUPE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183, US, 01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 6232456e  
US-08-944-483-35  
Alignment Scores:  
Pred. No.: 3,1e-30 Length: 224  
Score: 344.50 Matches: 76  
Percent Similarity: 54.25% Conservative: 39  
Best Local Similarity: 35.85% Mismatches: 94  
Query Match: 22.08% Indels: 3  
DB: 3 Gaps: 2  
US-10-037-270-482 (1-866) x US-08-944-483-35 (1-224)  
QY 108 AAGAAGACCTGCTGCTATTTGCTGACCTCAAGTCTCAACCTGCTGAGG 167  
DB 8 Glutlunserleuprotlyrglnvalserleuanserlgyserrhispecysglyly 27  
QY 168 GTCTCTCAACACCCAGCTGGGTCTGCCCCAGCTCACTGCTATTACCAATCTGAA 227  
DB 28 Serleuilesergluintrpvalvalseralalhisctyrylstrnrragllleqln 47  
QY 228 GTGATGCTGGGAATTTCAAGACAGAGCTCAAGACGCTACTCAAGACCAATTAACCC 287  
DB 48 Valargleuglygluhtlsaenilleysvalleugligngluignpheiieasmla 67

QY 288 ATTGAGATGCTCGCTACTGGAACCTACAGTCATAGGCCCCACAGATGACCTCATGCTC 347  
DB 68 Alalysillelearghisprolystryanarargasprthleuapnabpmlleu 87  
QY 348 ATCAAGCTGCTGAGCTGCTGATGCTCAATCCCAAGTCCAGCCCTCCCTCGGCACC 407  
DB 88 Iletyaleuserserproalvalilleasmlaargvalserthrlseerleuprotlr 107  
QY 408 ACCAATGTACGCCAGCAGCTGCTGCTACTCTGCTGAGTTTGACGTGAGCCAAAGAAC 467  
DB 108 Alaproprolalalaellythrgluysleuileseelityrpglysnthrlseurpne 127  
QY 468 AGTGCCGACACCTGACTGCGGCAACCTGAGACCCCTGATGCTATCGAGA 527  
DB 128 GlYAlasprytrproaprgluuyscybleuasplalprovalleuthrglnlaglu 147  
QY 528 TGCCAAAAACAGAACAGAAAACACACAGAAATTCCTATGTGTGAAATTTGGAAA 587  
DB 148 Cyslysalaserlyrprogllylsilthrnsenrmetphecysvalglphtleuglu 167  
QY 588 GTATTCAGCCGAATTTTGGGAGGTGCGCGTGTCTACTGATCTGCAAGACAGCTC 647  
DB 168 GLYGLYlyehapsercysgltnargapserglylyprovalvalcysanlglnleu 187  
QY 648 CAGGGA---ATCAGGTGGGACCTTCATGGA-----GGGACCTCGCATCTACCC 698  
DB 188 Glnglyvalvalsertrpghisgllysalatrplyasnarprprogllyvaltyrthr 207  
QY 699 AATGTTTCAATATGTATCTCGATTGAGAACACT 734  
DB 208 lyvaltyrasnlyrvalasprlleyasprthr 219  
RESULT 9  
US-08-956-267A-2  
Sequence 2, Application US/08956267A  
Patent No. 5945328  
GENERAL INFORMATION:  
APPLICANT: WOLDIKE, Helle Fabricius  
APPLICANT: KJELDSEN, Thomas Borglum  
TITLE OF INVENTION: A Process For Producing Trypsin  
TITLE OF INVENTION: (Trypsinogen)  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59453280 No. 5945328disk of No. 5945328th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,267A  
FILING DATE: 22-OCT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 4500, 204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 247 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-267A-2









DB 258 ProAspAlaProGlyValTyrThrLysValCysAsnTyrValAspTrpIleGlnAsnThr 277  
QY 735 GCTAAGAC 743  
DB 278 IleAlaAsp 280

RESULT 15  
US-08-278-091-9  
Sequence 9, Application US/08278091  
Patent No. 5506139  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: COWEN, Raymond P.  
APPLICANT: KLEIN, Michael H.  
TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with  
TITLE OF INVENTION: Reduced Protease Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/278, 091  
FILING DATE: 21-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-278-091-9

Alignment Scores:  
Pred. No.: 1,27e-26 Length: 223  
Score: 312.50 Matches: 71  
Percent Similarity: 52.66% Conservative: 38  
Best Local Similarity: 34.30% Mismatches: 95  
Query Match: 20.03% Indels: 3  
DB: 1 Gaps: 1

US-10-037-270-482 (1-866) x US-08-278-091-9 (1-223)

QY 123 CCTATTGTGTTACTCAAGTCTCACTCAACCCCTGTGTGGCGCTCTCATCAACC 182  
DB 13 PctyrglnAlaSerLeuAsnSerGlyTyrHisPheCysGlySerLeuIleAsnSer 32  
QY 183 AGCTGGTGTGCTGCCAGCTCACTGCTATTACCAATCTGAAGATGATGTGGAAAT 242  
DB 33 GlnTrpValValSerAlaAlaHisCysTyrLysSerGlyIleGlnValArgLeuGlyGln 52  
QY 243 TTCAGAGCAGAGTCAAGACGGTACTGAACAGACCAATTACCCCATTCAGATGTCGCC 302  
DB 53 AspAsnIleAsnValValGlnGlyAsnGlnGlnPheIleSerAlaSerLysSerIleVal 72

QY 303 TACTGGAAGTACATGATACGCCGCCACAGATGAGTCACTGATCAAGCTGAGCTAAG 362  
DB 73 HisProSerTyrAsnSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuLysSer 92  
QY 363 CCTGCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGCCACCAACCAATGTCAGCCA 422  
DB 93 AlaAlaSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAlaSerAla 112  
QY 423 GGCACGTGCTGCTACTCTCAGGTTTGAGCTGAGCCAGCAAGAAACAGTGGCCGACACCT 482  
DB 113 GlyThrGlnCysLeuIleSerGlyTyrGlyAsnThrLysSerSerGlyThrSerTyrPro 132  
QY 483 GACTTGGCGCAGAACCTTGAGGCCCCCTGATGTCGATGAGAAATGCCAAAAACAGAA 542  
DB 133 AspValIleuLysCysLeuLysAlaProIleLeuSerAspSerSerGlyLysSerAlaTyr 152  
QY 543 CAAGGAAAAAGCCACAGAAATTCCTTATGTGTGAATTGTGAAGTATTCAGCCCAATT 602  
DB 153 ProGlyGlnIleThrSerAsnMetPheCysAlaGlyTyrLeuGlnGlyLysAspSer 172  
QY 603 TTGGGAGGTGGCGCTGCTACTGTCATCTGCACCAAGACAGCTCCAGGAATC----- 656  
DB 173 CysGlnGlyAspSerGlyGlyProValValCysSerGlyLysLeuGlnGlyIleValSer 192  
QY 657 ---GAGGTGGGCACTTCATGGAGGGAGCTGCGCATCTACCAATGTTTACAATAT 713  
DB 193 TrpGlySerGlyCysAlaGlnLysAsnLysProGlyValTyrThrLysValCysAsnTyr 212  
QY 714 GATTCCTGATTCAGAAACT 734  
DB 213 ValSerTrpIleGlnThr 219

Search completed: November 8, 2003, 02:13:40  
Job time : 27 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 8, 2003, 02:11:47 ; Search time 181.5 Seconds  
(without alignments)  
1638.952 Million cell updates/sec

Title: US-10-037-270-482  
Perfect score: 1560  
Sequence: 1 ggcactactccctgcgtgcta.....tgggaatgcagcatcactagt 866

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 644079 segs, 171749292 residues

Total number of hits satisfying chosen parameters: 1288158

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=x1h  
-Q=/cgn2\_1/USFTO.epool/US10037270/runat.07112003.140520.8870/app\_query.fasta\_1.1031  
-DB=Published\_Applications\_AA -QFMT=fastan -SUFFI=rapb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.csi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10037270 @CGN\_1.1.50 @runat.07112003.140520.8870  
-NCPUS=6 -ICPU=3 -NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSPELCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match length	ID	Description
------------	-------	--------------------	----	-------------

1	1255	80.4	235	11	US-09-842-758-12	Sequence 12, App1
2	1255	80.4	235	11	US-09-842-758-14	Sequence 14, App1
3	1250	80.1	235	10	US-09-910-151-2	Sequence 2, App1
4	1247	79.9	235	12	US-10-246-006-2	Sequence 2, App1
5	1222	78.3	265	9	US-09-888-615-102	Sequence 102, App
6	1217	78.0	231	11	US-09-842-758-107	Sequence 107, App
7	377.5	24.2	248	11	US-09-842-758-54	Sequence 54, App1
8	375.5	24.1	248	11	US-09-842-758-53	Sequence 53, App1
9	351.5	22.5	246	11	US-09-842-758-55	Sequence 55, App1
10	334.5	21.4	247	9	US-09-923-779-154	Sequence 154, App
11	328	21.0	215	11	US-09-898-837A-30	Sequence 50, App1
12	319.5	20.5	281	14	US-10-021-368-7	Sequence 7, App1
13	312.5	20.0	223	10	US-09-910-071-14	Sequence 14, App1
14	311	19.9	243	12	US-10-239-663-39	Sequence 39, App1
15	305	19.6	241	12	US-10-137-870-248	Sequence 248, App
16	305	19.6	241	12	US-10-140-018-248	Sequence 248, App
17	305	19.6	241	12	US-10-140-021-248	Sequence 248, App
18	305	19.6	241	12	US-10-140-274-248	Sequence 248, App
19	305	19.6	241	12	US-10-140-471-248	Sequence 248, App
20	305	19.6	241	12	US-10-140-807-248	Sequence 248, App
21	305	19.6	241	12	US-10-140-922-248	Sequence 248, App
22	305	19.6	241	12	US-10-140-924-248	Sequence 248, App
23	305	19.6	241	12	US-10-140-926-248	Sequence 248, App
24	305	19.6	241	12	US-10-141-658-248	Sequence 248, App
25	305	19.6	241	12	US-10-141-702-248	Sequence 248, App
26	305	19.6	241	12	US-10-141-704-248	Sequence 248, App
27	305	19.6	241	12	US-10-143-432-248	Sequence 248, App
28	305	19.6	241	12	US-10-143-432-248	Sequence 248, App
29	305	19.6	241	12	US-10-143-767-248	Sequence 248, App
30	305	19.6	241	12	US-10-143-033-248	Sequence 248, App
31	305	19.6	241	12	US-10-144-994-248	Sequence 248, App
32	305	19.6	241	12	US-10-145-628-248	Sequence 248, App
33	305	19.6	241	12	US-10-145-631-248	Sequence 248, App
34	305	19.6	241	12	US-10-145-633-248	Sequence 248, App
35	305	19.6	241	12	US-10-145-746-248	Sequence 248, App
36	305	19.6	241	12	US-10-145-748-248	Sequence 248, App
37	305	19.6	241	12	US-10-145-823-248	Sequence 248, App
38	305	19.6	241	12	US-10-145-826-248	Sequence 248, App
39	305	19.6	241	12	US-10-145-870-248	Sequence 248, App
40	305	19.6	241	12	US-10-145-876-248	Sequence 248, App
41	305	19.6	241	12	US-10-145-959-248	Sequence 248, App
42	305	19.6	241	12	US-10-146-724-248	Sequence 248, App
43	305	19.6	241	12	US-10-146-725-248	Sequence 248, App
44	305	19.6	241	12	US-10-146-795-248	Sequence 248, App
45	305	19.6	241	12	US-10-147-435-248	Sequence 248, App

#### ALIGNMENTS

RESULT 1  
US-09-842-758-12  
; Sequence 12, Application US/09842758  
; Publication NO. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Vermet, Corine A. M.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Padigarat, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Gangoli, Reha A  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Rastelli, Luca  
; APPLICANT: MacDougall, John R  
; APPLICANT: Taupier, Raymond J  
; APPLICANT: Grose, William M

APPLICANT: Edward, Szekeres S  
APPLICANT: Aleobrook II, John P  
TITLE OF INVENTION: No. US20030083244A1e1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT FILING DATE: 2001-04-25  
CURRENT APPLICATION NUMBER: US/09/842,758  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,238  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 12  
LENGTH: 235  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-842-758-12  
Alignment Scores:  
Pred. No.: 2e-116 Length: 235  
Score: 1255.00 Matches: 235  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 80.45% Indels: 0  
DB: 11 Gaps: 0  
US-10-037-270-482 (1-866) x US-09-842-758-12 (1-235)  
QY 42 ATGAATATGCTCTTATTTGGGTGCTGCTGGGACATTTTCTTGTGACTACT 101  
DB 1 MetLeuTyrValPheTyrLeuGlyValLeuAlaGlyThrPhePheAlaAepSerSer 20  
QY 102 GTTCAGAAAGAAACCCCTGCTCCCTATTGCTGACTCAAGTCTCAACCCCTGT 161  
DB 21 ValGlnIleGlnAspProAlaProTyrLeuValTyrLeuIleSerHisPheAsnProCys 40  
QY 162 GTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCGCCAGCTCACTGCTATTACCAAT 221  
DB 41 ValGlyValLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 60  
QY 222 CTGAAGTGTGCTGGGAAATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 281  
DB 61 LeuIleValMetLeuGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIle 80  
QY 282 AACCCCTTGAATCGTCCCGCTACTGGAATCAAGTATAGCGCCCAAGATGACTC 341  
DB 81 AsnProIleGlnIleValIleIleIleIleIleIleIleIleIleIleIleIleIleIle 100  
QY 342 ATGCTCATCAAGCTGGCTAAGCTGACATGCTCAATCCCAAGTCCAGCCCTTCCCTC 401

DB 101 MetLeuIleIleValLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIle 120  
QY 402 GCCACACCAACCAATTCAGCCGACGACGACGACGACGACGACGACGACGACGACGAC 461  
DB 121 AlaThrThrAsnValAlaIleProGlyThrValCysLeuIleIleIleIleIleIleIleIle 140  
QY 462 GAAACAGTGGCCGACACCCCTGACTGCGGACAGAACCTGGAGGCGCCCGCTGATGTGAT 521  
DB 141 GluAsnSerIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 160  
QY 522 CGAAGATGCCAAAAACAGAACAGGAAAGCCACAGAGATTCCTTATGTGAAATTT 581  
DB 161 ArgGluCysGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 180  
QY 582 GTGAAGTATTCAGCCGAATTTTGGGGAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 641  
DB 181 ValIleValPheSerIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 200  
QY 642 AACCTCAGGGAATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 701  
DB 201 LysLeuGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 220  
QY 702 GTTACAAATATGTAATCTGGAATTGAGAACTGCTTAAGGACAG 746  
DB 221 ValTyrLeuTyrValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 235  
RESULT 2  
US-09-842-758-14  
Sequence 14, Application US/09842758  
Publication No. US20030083244A1  
GENERAL INFORMATION:  
APPLICANT: Vernet, Corine A. M.  
APPLICANT: Bernades, Elma R.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Shinkets, Richard A.  
APPLICANT: Malynkar, Uriel M.  
APPLICANT: Boldog, Ferenc L.  
APPLICANT: Zehusen, Bryan D.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Majumder, Komud  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Patirajan, Meera  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Gangoli, Beha A.  
APPLICANT: Smithson, Glenda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R.  
APPLICANT: Taupier, Raymond J.  
APPLICANT: Grose, William M.  
APPLICANT: Edward, Szekeres S  
APPLICANT: Aleobrook II, John P  
TITLE OF INVENTION: No. US20030083244A1e1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT APPLICATION NUMBER: US/09/842,758  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,238  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02

PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 14  
LENGTH: 235  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-842-758-14

Alignment Scores:  
Pred. No.: 2e-116 Length: 235  
Score: 1255.00 Matches: 235  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 80.45% Indels: 0  
DB: 11 Gaps: 0

US-10-037-270-482 (1-866) x US-09-842-758-14 (1-235)

QY 42 ATGAATATGCTCTTATTTGGGTGCTCGCTGGAGACATTTTCTTGGTCACTATCT 101  
DB 1 MetUyTyValPheTyTLeuGlyValLeuAlaGlyThrPhePheAlaPheSerSer 20  
QY 102 GTTCAGAAAGAGACCTGCTCTCTTATTTGGTGTACCTCAAGTCACTTCAACCCCTGT 161  
DB 21 ValGlnLysGlnAspProAlaProTyLeuValTyLeuLysSerHisPheAsnProCys 40  
QY 162 GTGGGGCTCTCTATCAAAACCCAGCTGGGTGCTGGCCCACTCACTGCTATTTACCAAT 221  
DB 41 ValGlyValLeuLysProSerTrpValLeuAlaProAlaHisCysTyLeuProAsn 60  
QY 222 CTGAAGTGTGCTGGAAATTTCAAGACAGAGTCAAGACGTACTGAACAGACAAT 281  
DB 61 LeuLysValMetLeuGlnAsnPheLysSerArgValArgAspGlyThrGlnIntrHis 80  
QY 282 AACCCCATTCAGATGCTCGCTACTGAATCAAGTCAATAGCCGCCACAGATGACCTC 341  
DB 81 AsnProLysGlnLeuValArgTyTrpAsnTySerHisSerAlaProGlnAspLeu 100  
QY 342 ATGCTATCAAGCTGGCTAAGCTGCTGATGCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
DB 101 MetLeuLysLeuValLysProAlaMetLeuAsnProLysValGlnProLeuProLeu 120  
QY 402 GCCACCAACCAATGTGAGCCAGGACATGCTGTCTACTCTCAAGTTTGGACTGGAGCAA 461  
DB 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140  
QY 462 GAAACAGTGGCCGACACCTGACTTGGCGAGAACCTGAGAGCCGCCGTGATGTCTGAT 521  
DB 141 GlnAsnSerGlyArgHisProAspLeuArgGlnAsnLeuGlnAlaProValMetSerAsp 160  
QY 522 CGAGAAATGCCAAAAACAGAACAGAAAAAGCCACAGAAATCTTATGTGTAATTT 581  
DB 161 ArgGlnCysGlnLysThrGlnGlnGlyLysSerHisArgAsnSerLeuCysValLysPhe 180  
QY 582 GTGAAGATTTTACCGCAATTTTGGGAGGTGGCGCTGCTACTGCTCAATCGCAAGAC 641  
DB 181 ValLysValPheSerArgHisPheGlyGlnValAlaValAlaThrValLysCysLysAsp 200  
QY 642 AACCTCAGGAAATCAGATGAGGCACTTCAATGAGAGGAGCGCTGCACTTCAACAAT 701  
DB 201 LysLeuGlnGlnGlyLysGlnValGlnHisPheMetCysGlyAspValGlyLysLeuThrAsn 220

QY 702 GTTACAAATATGATCTCTGATGAGAACCTGCTAAGACAAG 746  
DB 221 ValTyTyTyValPheSerTrpLeuGlnAsnThrAlaLysAspLys 235

RESULT 3  
US-09-910-151-2  
Sequence 2, Application US/09910151  
Patent No. US20020137181A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 14087, A NOVEL SERINE PROTEASE MOLECULE  
TITLE OF INVENTION: AND USES THEREFOR  
FILE REFERENCE: 38155-20021.00  
CURRENT APPLICATION NUMBER: US/09/910,151  
CURRENT FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US 60/219,002  
PRIOR FILING DATE: 2000-07-17  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 235  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-910-151-2

Alignment Scores:  
Pred. No.: 6.31e-116 Length: 235  
Score: 1250.00 Matches: 234  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.57% Mismatches: 0  
Query Match: 80.13% Indels: 0  
DB: 10 Gaps: 0

US-10-037-270-482 (1-866) x US-09-910-151-2 (1-235)

QY 42 ATGAATATGCTCTTATTTGGGTGCTCGCTGGAGACATTTTCTTGGTCACTATCT 101  
DB 1 MetUyTyValPheTyTLeuGlyValLeuAlaGlyThrPhePheAlaPheSerSer 20  
QY 102 GTTCAGAAAGAGACCTGCTCTCTTATTTGGTGTACCTCAAGTCACTTCAACCCCTGT 161  
DB 21 ValGlnLysGlnAspProAlaProTyLeuValTyLeuLysSerHisPheAsnProCys 40  
QY 162 GTGGGGCTCTCTATCAAAACCCAGCTGGGTGCTGGCCCACTCACTGCTATTTACCAAT 221  
DB 41 ValGlyValLeuLysProSerTrpValLeuAlaProAlaHisCysTyLeuProAsn 60  
QY 222 CTGAAGTGTGCTGGAAATTTCAAGACAGAGTCAAGACGTACTGAACAGACAAT 281  
DB 61 LeuLysValMetLeuGlnAsnPheLysSerArgValArgAspGlyThrGlnIntrHis 80  
QY 282 AACCCCATTCAGATGCTCGCTACTGAATCAAGTCAATAGCCGCCACAGATGACCTC 341  
DB 81 AsnProLysGlnLeuValArgTyTrpAsnTySerHisSerAlaProGlnAspLeu 100  
QY 342 ATGCTATCAAGCTGGCTAAGCTGCTGATGCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
DB 101 MetLeuLysLeuValLysProAlaMetLeuAsnProLysValGlnProLeuProLeu 120  
QY 402 GCCACCAACCAATGTGAGCCAGGACATGCTGTCTACTCTCAAGTTTGGACTGGAGCAA 461  
DB 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140  
QY 462 GAAACAGTGGCCGACACCTGACTTGGCGAGAACCTGAGAGCCGCCGTGATGTCTGAT 521  
DB 141 GlnAsnSerGlyArgHisProAspLeuArgGlnAsnLeuGlnAlaProValMetSerAsp 160  
QY 522 CGAGAAATGCCAAAAACAGAACAGAAAAAGCCACAGAAATCTTATGTGTAATTT 581  
DB 161 ArgGlnCysGlnLysThrGlnGlnGlyLysSerHisArgAsnSerLeuCysValLysPhe 180  
QY 582 GTGAAGATTTTACCGCAATTTTGGGAGGTGGCGCTGCTACTGCTCAATCGCAAGAC 641

```
Db 181 VallyValPheSerArgIlePheGlyValValAlaValAlaThrValIleCysLysAsp 200
Qy 642 AAGTCAGGAATTCAGAGTGGGCACTTTCATGGAGGAGGAGCGGCACTTACCAAT 701
Db 201 LysLeuGlnGlyIleGlyValGlyHisPheMetGlyGlyAspValGlyIleTyThrAsn 220
Qy 702 GTTACAAATATGATCTCGATTTGAGAACACTGCTAAGGACAA 746
Db 221 ValTyIleTyValSerTrpIleGluAsnThrAlaLysAspLys 235

RESULT 4
US-10-246-006-2
; Sequence 2, Application US/10246006
; Publication No. US2003014493A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; TITLE OF INVENTION: Human Serine Protease
; FILE REFERENCE: 00-16
; CURRENT APPLICATION NUMBER: US/10/246,006
; CURRENT FILING DATE: 2002-09-17
; NUMBER OF SEQ. ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-246-006-2

Alignment Scores:
Pred. No.: 1,266-115 Length: 235
Score: 1247.00 Matches: 234
Percent Similarity: 99.57% Conservative: 0
Best Local Similarity: 99.57% Mismatches: 1
Query Match: 79.94% Indels: 0
DB: 12 Gaps: 0

US-10-037-270-482 (1-866) x US-10-246-006-2 (1-235)
Qy 42 ATGAAATATGCTCTTCTATTGGGCTGCTGCTGGAGCAATTTTCTTGCTGACTATCT 101
Db 1 MetLysTyValPheTyLeuGlyValLeuAlaGlyThrPhePhePheAlaAspSerSer 20
Qy 102 GTTCAGAAAGAAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 161
Db 21 ValGlnLysGlnAspProAlaProTyLeuValTyLeuLysSerHisPheAsnProCys 40
Qy 162 GTGGGCTGCTCATCAACCCAGCTGGGCTGGCCCAAGCTCACTCTATTACCAAT 221
Db 41 ValGlyValLeuIleLysProSerTrpValLeuAlaProAlaHisCysTyLeuProAsn 60
Qy 222 CTGAAATGATGCTGGGAAATTTTCAAGCAGACTCAGACGCTATCAAGACCAAT 281
Db 61 LeuLysValMetLeuGlyAsnPheLysSerArgValArgAspGlyThrGluGlnThrIle 80
Qy 282 AACCCATTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 341
Db 81 AsnProIleGlnIleValArgTyTrpAsnTySerHisSerAlaProGlnAspAspLeu 100
Qy 342 ATGCTCATCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401
Db 101 MetLeuIleLysLeuAlaLysProAlaMetLeuAsnProLysValGlnProLeuThrLeu 120
Qy 402 GCCACCAATGCTCAGGCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461
Db 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140
Qy 462 GAAAGAGTGGCGGACACCTGATTCGGGAGAACTGAGAGGCGCCCGGATGCTGAT 521
Db 141 GluAsnSerGlyArgHisProAspLeuArgGlnAsnLeuGlnAlaProValMetSerAsp 160
Qy 522 CGAGAAATGCCAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATTCCTATGCTGAATTT 581
```

```
Db 161 ArgGluCysGlnLysThrGlnGlnGlyLysSerHisArgAsnSerLeuCysValLysPhe 180
Qy 582 GTGAAGTATTCAGCCGAATTTTGGGAGAGTGGCCGCTGCTGCTGCTGCTGCTGCTGCT 641
Db 181 ValLysValPheSerArgIlePheGlyValAlaValAlaThrValIleCysLysAsp 200
Qy 642 AAGTCAGGAATTCAGAGTGGGCACTTTCATGGAGGAGGAGCGGCACTTACCAAT 701
Db 201 LysLeuGlnGlyIleGlyValGlyHisPheMetGlyGlyAspValGlyIleTyThrAsn 220
Qy 702 GTTACAAATATGATCTCGATTTGAGAACACTGCTAAGGACAA 746
Db 221 ValTyIleTyValSerTrpIleGluAsnThrAlaLysAspLys 235

RESULT 5
US-09-888-615-102
; Sequence 102, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CHENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ. ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 102
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-102

Alignment Scores:
Pred. No.: 4,096-113 Length: 265
Score: 1222.00 Matches: 234
Percent Similarity: 88.30% Conservative: 0
Best Local Similarity: 88.30% Mismatches: 1
Query Match: 78.33% Indels: 30
DB: 9 Gaps: 1

US-10-037-270-482 (1-866) x US-09-888-615-102 (1-265)
Qy 42 ATGAAATATGCTCTTCTATTGGGCTGCTGCTGGAGCAATTTTCTTGCTGACTATCT 101
Db 1 MetLysTyValPheTyLeuGlyValLeuAlaGlyThrPhePhePheAlaAspSerSer 20
Qy 102 GTTCAGAAAGAAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 161
Db 21 ValGlnLysGlnAspProAlaProTyLeuValTyLeuLysSerHisPheAsnProCys 40
Qy 162 GTGGGCTGCTCATCAACCCAGCTGGGCTGGCCCAAGCTCACTCTATTACCAAT 221
Db 41 ValGlyValLeuIleLysProSerTrpValLeuAlaProAlaHisCysTyLeuProAsn 60
Qy 222 CTGAAATGATGCTGGGAAATTTTCAAGCAGACTCAGACGCTATCAAGACCAAT 281
Db 61 LeuLysValMetLeuGlyAsnPheLysSerArgValArgAspGlyThrGluGlnThrIle 80
Qy 282 AACCCATTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 341
Db 81 AsnProIleGlnIleValArgTyTrpAsnTySerHisSerAlaProGlnAspAspLeu 100
Qy 342 ATGCTCATCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401
Db 101 MetLeuIleLysLeuAlaLysProAlaMetLeuAsnProLysValGlnProLeuThrLeu 120
```

QY 402 GCCACCAACCAATGTCAGGCGGCACTGTCTGCTACTCTCAGGTTTGAAGTGAAGCAA 461  
| | | | |  
Db 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140  
| | | | |  
QY 462 GAAACACGT----- 470  
| | | | |  
Db 141 GluAsnSerGlyLeuTrpGlnLeuGluProProGlyHisLeuThrLeuHisArgGlyPro 160  
| | | | |  
QY 471 -----GCGCGACACCTGACCTGGCGG 491  
| | | | |  
Db 161 AlaIleProAspTrpGlnArgHisAsnSerHisGluGlnIleArgHisProAspLeuArg 180  
| | | | |  
QY 492 CAGAACCTGAGGCGCCCGGATGCTGATCGAAGATGCCAAAAACAGAAACAGAGAAA 551  
| | | | |  
Db 181 GluAsnLeuGlnIleProValMetSerAspArgGluGlnSerGlnThrGluGlnGlyLe 200  
| | | | |  
QY 552 AGCCACGAGAAATCTTATGTGTGAATTTGTGAAGTATTCAGCCGAATTTTGGGAG 611  
| | | | |  
Db 201 SerHisArgAsnSerLeuGlyValValPheValValPheSerArgIlePheGlyGlu 220  
| | | | |  
QY 612 GTGCGCGTCTCTGCTGATCTGCAAGACAGAGCTCCAGGAGTCGAGGCGGCACTTC 671  
| | | | |  
Db 221 ValAlaValAlaThrValIleCysLysAspLysLeuGlnIleGluValGlyHisPhe 240  
| | | | |  
QY 672 ATGGAGGGGAGCGTGGCATCTACACCAATGTTTACAAATATGATCTGATTTGAGAAC 731  
| | | | |  
Db 241 MetGlyGlyAspValGlyIleTyrrThrAsnValTyrrLysTyrrValSerTrpIleGluAsn 260  
| | | | |  
QY 732 ACTGCTTAAGACAAAG 746  
| | | | |  
Db 261 ThrAlaLysAspLys 265  
| | | | |

## RESULT 6

US-09-842-758-107  
; Sequence 107, Application US/09842758  
; Publication No. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Fernandez, Elma R.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Zehusen, Bryan D.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Majumder, Kumar T.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patnirajan, Meera  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Gangoli, Bha A.  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Rastelli, Luca  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Taupier, Raymond J.  
; APPLICANT: Grose, William M.  
; APPLICANT: Edward, Szekeres S.  
; APPLICANT: Alsobrook II, John P.  
; TITLE OF INVENTION: No. US20030083244A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-783  
; CURRENT APPLICATION NUMBER: US/09/842,758  
; CURRENT FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/200,158  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: 60/200,613  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,780  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/201,006  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,007  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,236

; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,238  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,186  
; PRIOR FILING DATE: 2000-05-02  
; PRIOR APPLICATION NUMBER: 60/201,474  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 60/201,508  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 60/220,591  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/232,678  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 60/263,217  
; PRIOR FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: 60/265,160  
; PRIOR FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 107  
; LENGTH: 231  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-842-758-107  
Alignment Scores:  
Score: 1,246-112 Length: 231  
Percent Similarity: 127.00 Matches: 227  
Best Local Similarity: 98.704 Conservative: 0  
Query Match: 98.704 Mismatches: 3  
Query Match: 78.014 Indels: 1  
DB: 11 Gaps: 0

US-10-037-270-482 (1-866) x US-09-842-758-107 (1-231)

QY 42 ATGAATATGCTCTTCTATTGGGTGCTCTGCGGACATTTTCTTGTGACTCATCT 101  
| | | | |  
Db 1 MetLysTyrrValPheTyrrLeuGlyValLeuAlaGlyThrPhePheHehAlaAspSerSer 20  
| | | | |  
QY 102 GTTCAGAAAGAAAGACCTGCTGCTCTTATTGGGTGACTCAAGTCTCACTTCAACCCCTGT 161  
| | | | |  
Db 21 ValGlnLysGlyAspProAlaProTyrrLeuValTyrrLeuLysSerHisPheAsnProCys 40  
| | | | |  
QY 162 GTGGCGGCTCTATCAACCCAGCTGGGTGCGGCCCGCCAGCTCATATTATACCAAT 221  
| | | | |  
Db 41 ValGlyValLeuIleLysProSerTrpValLeuAlaProAlaHisCysTyrrLeuProAsn 60  
| | | | |  
QY 222 CTGAAGATGATGCTGGGAATTTCAAGACAGAGTCAAGACGCTGTAACAGCAATT 281  
| | | | |  
Db 61 LeuLysValMetLeuGlnAsnPheLysSerArgValArgAspGlyThrGluGlnTrpIle 80  
| | | | |  
QY 282 AACCCCATTCAGATGCTCGCTACTGGAATCAAGTCAAGTACGAGCCGACAGATGACCTC 341  
| | | | |  
Db 81 AsnProIleGlnIleValArgTyrrTrpAsnTyrrSerHisSerAlaProGlnAspLeu 100  
| | | | |  
QY 342 ATGCTATCAAGCTGCTTAAGCTGCGATGCTCAATCCAAAGTCCAGGCCCTTCCCCTC 401  
| | | | |  
Db 101 MetLeuIleLysLeuAlaLysProAlaMetLeuAsnProLysValGlnProLeuThrLeu 120  
| | | | |  
QY 402 GCCACCAACCAATGTCAGGCGGCACTGTCTGCTACTCTCAGGTTTGAAGTGAAGCAA 461  
| | | | |  
Db 121 AlaThrThrAsnValArgProGlyThrValCysLeuLeuSerGlyLeuAspTrpSerGln 140  
| | | | |  
QY 462 GAAACACGTGGCGGACCCCTGACTTGGCGGAGAACCTGAGTGTGTGAT 521  
| | | | |  
Db 141 GluAsnSerGlyArgHisProAspLeuArgIleAsnLeuGlnAlaProValMetSerAsp 160  
| | | | |  
QY 522 CAGAAATGCCA-AAAAACGAACAGAGAAAAAGCCACAGAAATTCCTTATGTGTGAATT 580  
| | | | |  
Db 161 ArgGluCysGlnLysAsnArgThrArgLysProGlnGlnLysPheLeuMetCysGluIle 180  
| | | | |  
QY 581 TGTGAAGTATTCAGGCGGAAATTTTGGGAGGCGGCTTGGTACTGTCATCTGACAGAGA 640  
| | | | |

Db 181 CysGluSerIleGlnProAsnPhnTrpGlyGlyAlaArgCysTyrCysHisLeuGlnArg 200  
Qy 641 CAAGCTCCAGGAGATCGAGGTGGGACCTTCATGAGGAGGAGCGTCCGACATCTACACCAA 700  
Db 201 GlnAlaProGlnAsnAlaGlyGlyAlaLeuHisGlyArgGlyAlaGlnHisLeuHisGln 220  
Qy 701 TGTTTACAAATATGTATTCCTGGATTGACA 730  
Db 221 CysLeuGlnIleCysIleLeuAspSerGlu 230

## RESULT 7

US-09-842-758-54  
; Sequence 54, Application US/09842758  
; Publication No. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Verneet, Corine A. M.  
; APPLICANT: Fernandez, Elma R.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Majumder, Kunud T.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Paturajan, Meera  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Gangolli, Beba A.  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Rastelli, Luca  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Taupier, Raymond J.  
; APPLICANT: Grose, William M.  
; APPLICANT: Edwards, Szekeres S.  
; APPLICANT: Alsobrook II, John P.  
; TITLE OF INVENTION: No. US20030083244A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-783  
; CURRENT APPLICATION NUMBER: US/09/842,758  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/200,158  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: 60/200,613  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,780  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/201,006  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,007  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,236  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,238  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,186  
; PRIOR FILING DATE: 2000-05-02  
; PRIOR APPLICATION NUMBER: 60/201,474  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 60/201,508  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 60/220,591  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/232,678  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 60/263,217  
; PRIOR FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: 60/265,160  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54  
; LENGTH: 248  
; TYPE: PRT

; ORGANISM: Gallus gallus  
US-09-842-758-54  
Alignment Scores:  
Pred. No.: 86-29 Length: 248  
Score: 377.50 Matches: 88  
Percent Similarity: 53.06% Conservative: 42  
Best Local Similarity: 35.92% Mismatches: 100  
Query Match: 24.20% Gaps: 15  
DB: 11 Gaps: 4

US-10-037-270-482 (1-866) x US-09-842-758-54 (1-248)

Qy 42 ATGAATAT-----GCTTCTATTGGGTGCTGCTGGAATTTTCTTGGCT 92  
Db 1 MetLysPheLeuValLeuValAlaPheLeuGlyVal---AlaValAlaPheProIleSer 19  
Qy 93 GACTCATCTGTTAGAAA-----GAAACCTGCTCCCTAT 128  
Db 20 AspGluAspAspAspIleValGlyGlyTyrSerCysAlaArgSerAlaIleProTyr 39  
Qy 129 TTGGTGTACTCAAGTCTCACTCAACCCCTGTGTGGGCGCTCTATCAACCCAGCTG 188  
Db 40 GlnValSerLeuAsnSerGlyTyrHisPheCysGlyGlySerLeuIleSerSerGlnTyr 59  
Qy 189 GTCTGGCCCAAGCTCACTGCTATTATCAAAATCTGAAGTGAATGCTGGGAATTTCAAG 248  
Db 60 ValLeuSerAlaAlaHisCysTyrTyrSerSerIleGlnValLeuGlyGlyTyrAsn 79  
Qy 249 AGCAGAGTCAGACGGTACTGAACAGCAATTAACCCATTGATCGTCCGCTACTGG 308  
Db 80 LeuAlaGlnAspIleSerGlnGlnThrIleSerSerSerValIleArgHisSer 99  
Qy 309 AACTACAGTATGAGCCGCCCAAGATGACCTGATCTATCAAGCTGAGTACCTGCC 368  
Db 100 GlyTyrAsnSerAsnThrLeuAsnAspIleMetLeuIleLysLeuSerIleAlaAla 119  
Qy 369 ATCTCAATCCCAAGATCCAGCCCTTCCCTCGCCACCAACCAATGTCAAGCCAGCCT 428  
Db 120 ThrLeuAsnSerTyrValAsnThrValProLeuProThrSerCysValThrAlaGlyThr 139  
Qy 429 GTCTGTCTACTCTCAGGTTTGACCTGAGCCAAAGAAACAGTGGCCGACCTGACTTG 488  
Db 140 ThrCysLeuIleSerGlyTyrGlyAsnThrLeuSerSerGlySerLeuTyrProAspVal 159  
Qy 489 CGGCAGAACCTGAGGCCCCCGATGTCGATCGAATGCCAATAAACAGAAACAGAG 548  
Db 160 LeuGlnCysLeuAsnAlaProValLeuSerSerSerGlnCysSerSerAlaTyrProGly 179  
Qy 549 AAAGCCACAGAAATCTTATGTGAAATTTGTGAAGTATTCAGCCGAATTTTGGG 608  
Db 180 ArgIleThrSerAsnMetIleCysIleGlyTyrLeuAsnGlyGlyIleAspSerCysGln 199  
Qy 609 GAGTGGCCGTTGCTACTGCTATCTCAAAAGACACTCCAGGAATC-----GAG 659  
Db 200 GlyAspSerGlyGlyProValValCysAsnGlyGlnLeuGlnGlyIleValSerTrpGly 219  
Qy 660 GTGGGACACTTCATGAGGAGCGTCGGCATCTACCAATGTTTAAATATGATATCC 719  
Db 220 IleGlyCysAlaGlnLysGlyTyrProGlyValTyrThrLysValCysAsnTyrValSer 239  
Qy 720 TGGATTGAGAACT 734  
Db 240 TrpIleLysThrThr 244

RESULT 8  
US-09-842-758-53  
; Sequence 53, Application US/09842758  
; Publication No. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Verneet, Corine A. M.  
; APPLICANT: Fernandez, Elma R.  
; APPLICANT: Gerlach, Valerie

APPLICANT: Shimkets, Richard A  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Zerhusen, Bryan D  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Paturajan, Meera  
APPLICANT: Burgess, Catherine E  
APPLICANT: Gangoli, Beha A  
APPLICANT: Smithson, Glenda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R  
APPLICANT: Taupier, Raymond J  
APPLICANT: Groese, William M  
APPLICANT: Edward, Szekeres S  
APPLICANT: Alsobrook II, John P  
FILE OF INVENTION: No. US20030083244A1e1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT APPLICATION NUMBER: US/09/842,758  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 53  
LENGTH: 248  
TYPE: PRF  
ORGANISM: Gallus gallus  
US-09-842-758-53  
Alignment Scores:  
Pred. No.: 1,27e-28 Length: 248  
Score: 375.50 Matches: 87  
Percent Similarity: 53.06% Conservative: 43  
Best Local Similarity: 35.51% Mismatches: 100  
Query Match: 24.07% Indels: 15  
DB: 11 Gaps: 4  
US-10-037-270-482 (1-866) x US-09-842-758-53 (1-248)  
QY 42 ATCAATAT-----GCTCTATTTGGGTCTCTCGTGGACATTTTCTTGTCT 92  
DB 1 MeLVpPheuvValleuValAlaPheValIglyValThrVal---AlaPheProIleSer 19  
QY 93 GACTCATCTGTTCAGAAA-----GAAGACCTGTCTCCTAT 128

DB 20 AaPGLuAspAspAspLysLeValIglyIYrSerCyAlaArgSerAlaIaIaProIYr 39  
QY 129 TTGGTGTACTACCTGAAGCTCATTCAACCCCTGTGGGGTCTCTATCAACCCAGCTG 188  
DB 40 GlnValSerLeuAsnSerIYrThrSpheCySgIYrSerLeuIleSerSerGlnTrp 59  
QY 189 GTGCTGGCCCCAGCTCATTCTATTACCAATCTGAAGTGAATGCTGGAAATTTCAAG 248  
DB 60 ValLeuSerAlaAlaHisCyStrIYrIYrSerSerIleGlnValIYrLeuIYrIYrAsn 79  
QY 249 AGCAGAGTCAGAGCCGATCTGAACAGACATTTAACCCATTGATCGCTGCTG 308  
DB 80 LeuAlaIaIaGlnAspIYrSerGlnGlnTrIleSerSerSerIYrValIleArgHisSer 99  
QY 309 AACTACAGTCATAGCCCCCAAGATGACCTTCATGCTATCAAGCTGCTGCTGCC 368  
DB 100 GILYTYrAsnAlaAsnThrLeuAsnAsnAspIleMetLeuIleYrLeuSerIYrAlaAla 119  
QY 369 ATGCTCAATCCCAAGTCCAGCCCTTCCCTGCGCACCAATGTCAGGCCAGGCTGCT 428  
DB 120 ThrLeuAsnSerTYrValAsnThrValProLeuProThrSerCyValThrAlaGlyThr 139  
QY 429 GTCTGTCTACTCTCAGGTTTGAATGACCTGACCAAGAAAGATGCGGACACCTGACT 488  
DB 140 ThrCyLeuIleSerGlyTrpGlyAsnThrLeuSerSerGlySerLeuTYrProAspVal 159  
QY 489 CCGCAGAACCTGAGAGCCCCCTGATGCTGTGATCGAAATGCCAAAACAGAACAGAA 548  
DB 160 LeuGlnCyLeuAsnAlaProValLeuSerSerSerGlnCySerSerAlaTYrProGly 179  
QY 549 AAAGCCACAGAAATCTTATGTGTGAATTTGTGAAGTATTCAGCCGAATTTTGG 608  
DB 180 ArgIleThrSerAsnMetIleCyStrIleGlyTYrLeuAsnGlyIYrYrAspSerCyGln 199  
QY 609 GAGTGGCCGTTGCTACTGTCACTGCAATGCAAGCAAGCTCAGGAATC-----GAG 659  
DB 200 GLYAspSerGlyIYrProValValCyAsnGlnIYrLeuGlnIYrIleValSerTrpGly 219  
QY 660 GTGGGCACTTATGAGGAGGAGCGCGCATCTACCAATGTTCAAAATTTATATC 719  
DB 220 IlegIYrCyAlaGlnIYrSerIYrProGlyValTYrThrIYrValCyAsnTYrValSer 239  
QY 720 TGGATTGAGACACT 734  
DB 240 TrpIleYrThrThr 244  
RESULT 9  
US-09-842-758-55  
Sequence 55, Application US/09842758  
Publication No. US20030083244A1  
GENERAL INFORMATION:  
APPLICANT: Vernet, Corine A. M.  
APPLICANT: Fernandes, Elma R.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Shimkets, Richard A  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Zerhusen, Bryan D  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Paturajan, Meera  
APPLICANT: Burgess, Catherine E  
APPLICANT: Gangoli, Beha A  
APPLICANT: Smithson, Glenda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R  
APPLICANT: Taupier, Raymond J  
APPLICANT: Groese, William M  
APPLICANT: Edward, Szekeres S  
APPLICANT: Alsobrook II, John P



TITLE OF INVENTION: No. US20030083244A1el Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT APPLICATION NUMBER: US/09/842,758  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,238  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 55  
LENGTH: 246  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-842-758-55

Alignment Scores:  
Pred. No.: 3,14e-26 Length: 246  
Score: 351.50 Matches: 77  
Percent Similarity: 53.49% Conservative: 38  
Best Local Similarity: 35.81% Mismatches: 97  
Query Match: 22.53% Indels: 3  
DB: 11 Gaps: 1

US-10-037-270-482 (1-866) x US-09-842-758-55 (1-246)

QY 108 AAGAAGACCTGCTGCTATTTGGTGTACCTCAAGCTCACTTCAACCCCTGTGGGC 167  
DB 31 ArgGluSerSerValProTyrGlnValSerLeuAsnAlaGlyTyrHisPheCysGlyGly 50  
QY 168 GTCCTCATCAAAACCCAGCTGGGTGGTGGCCCACTCACTGATTTACCAATTCGA 227  
DB 51 SerLeuHisAsnArgGlnTrpValValSerAlaAlaHisCysGlyTyrLysTyrArgGln 70  
QY 228 GTGATCTGGGAATTTCAAGACAGCTCAGACGGTATCGAAGACAAATTAACCC 287  
DB 71 ValArgLeuGlyGlnHisAsnHisLeuValLeuGlnGlyAsnGlnInPheValAspSer 90  
QY 288 ATTCAGATGCTGGCTAGTGAAGTCAAGTCAAGGAGCCCAAGGATGACCTCATGCTC 347  
DB 91 AlaValIleIleLeuArgHisProAsnTyrAsnSerTrpThrLeuAspAsnSerIleLeu 110  
QY 348 ATCAAGCTGCTAGGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTGGCACC 407  
DB 111 IleValLeuAlaSerProValThrLeuAsnAlaArgValAlaSerValProLeuProSer 130  
QY 408 ACCAATGTCAAGGCAAGCACTGTCTGTCTTCTGAGTTTGAATGAGCAAGCAAAAC 467  
DB 131 SerCysAlaProAlaGlyThrGlnCysLeuIleSerGlyTyrPglYasnThrLeuSerAsn 150

QY 468 AGTGGCCGACACCTGCTGCGGAGAACTTGAGAGCCCTGATGTCGATCGAGA 527  
DB 151 GlyValAsnAsnProAsnProLeuGlnCysValAspAlaProValLeuProGlnAlaAsp 170  
QY 528 TGGCAAAAACAGAACAGAAAAGCCACAGGAATTCCTTATGTGTGAATTTGGAAA 587  
DB 171 CysGluAlaSerTyrProGlyAspIleThrAsnAsnMetIleCysValGlyPheLeuGlu 190  
QY 588 GTATTCAGCCGAATTTTGGGAGAGTGGCCGCTGCTACTGCTCATCTGCAAGCAAGCTC 647  
DB 191 GlyGlyLysAspSerCysGlnGlyAspSerGlyGlyProValValCysAsnGlyGluLeu 210  
QY 648 CAGGGAATC-----GAGTGGGCACTTCATGAGAGGGGAGCGTCGATCAACC 698  
DB 211 GlnGlyIleValSerTrpGlyTyrGlyCysAlaGlnProAspAlaProGlyValTyrThr 230  
QY 699 AATGTTTACAAATATGATTCCTGATTTGAGAACACTGCTTAAGAC 743  
DB 231 LysValCysAsnTyrValAspTrpIleGlnAsnThrIleAlaAsp 245

RESULT 10

US-09-923-779-154  
Sequence 154, Application US/09923779  
Patent No. US20020076721A1  
GENERAL INFORMATION:  
APPLICANT: Pyle, Ruth A.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Kaios, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.553  
CURRENT APPLICATION NUMBER: US/09/923,779  
NUMBER OF SEQ ID NOS: 155  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 154  
LENGTH: 247  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-923-779-154

Alignment Scores:  
Pred. No.: 1.57e-24 Length: 247  
Score: 334.50 Matches: 75  
Percent Similarity: 52.83% Conservative: 37  
Best Local Similarity: 35.38% Mismatches: 97  
Query Match: 21.44% Indels: 3  
DB: 9 Gaps: 1

US-10-037-270-482 (1-866) x US-09-923-779-154 (1-247)

QY 108 AAGAAGACCTGCTGCTATTTGGTGTACCTCAAGCTCACTTCAACCCCTGTGGGC 167  
DB 31 GluGluAsnSerValProTyrGlnValSerLeuAsnSerGlyTyrHisPheCysGlyGly 50  
QY 168 GTCCTCATCAAAACCCAGCTGGGTGGTGGCCCACTCACTGATTTACCAATTCGA 227  
DB 51 SerLeuHisSerGlnGlnTrpValValSerAlaAlaGlyHisCysGlyTyrLysSerArgIleGln 70  
QY 228 GTGATCTGGGAATTTCAAGACAGCTCAGACGGTATCGAAGACAAATTAACCC 287  
DB 71 ValArgLeuGlyGlnHisAsnHisLeuValLeuGlnGlyAsnGlnInPheValAspSer 90  
QY 288 ATTCAGATGCTGGCTAGTGAAGTCAAGTCAAGGAGCCCAAGGATGACCTCATGCTC 347  
DB 91 AlaValIleIleLeuArgHisProLysTyrAsnSerArgThrLeuAspAsnSerIleLeu 110  
QY 348 ATCAAGCTGCTAGGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTGGCACC 407  
DB 111 IleValLeuSerSerProAlaValIleAsnSerArgValSerAlaIleSerLeuProThr 130  
QY 408 ACCAATGTCAAGGCAAGCACTGTCTGTCTTCTGAGTTTGAATGAGCAAGCAAAAC 467



```

Db      131 AlaProAlaAlaGlyThrGluSerLeuIleSerGlyTrpGlyAsnThrLeuSerSer 150
Qy      468 AGTGGCCGACACCTGACTCTGGCGGAGAACCTGGAGAGCCCGCTGATGCTGATGACAGA 527
Db      151 G1yAlaAspTrpProAspGluLeuGlnCysLeuAspAlaProValLeuSerGlnAlaGlu 170
Qy      528 TGGCAAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 587
Db      171 CysGluAlaSerTrpProGlyLysIleThrAsnAspMetPheCysValGlyIleLeuGln 190
Qy      568 GTATTCAAGCGAATTTTGGGAGGTGGCGCTGCTACTGATGATGATGATGATGATGATGATG 647
Db      191 G1yGlyLysAspSerCysGlnGlyAspSerGlyGlyProValValSerAsnGlyIleLeu 210
Qy      648 CAGGGAATC-----GAGTGGGGGCACTTCATGGAGGGAGGAGGAGGAGGAGGAGGAGGAG 698
Db      211 G1nglyIleValaSerTrpGlyTrpGlyCysAlaGlnLysAsnArgProGlyValaLys 230
Qy      699 AATGTTCAATATGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 734
Db      231 LysValLysAsnTrpValaSerTrpIleLysAspThr 242

```

## RESULT 11

US-09-898-837A-50  
 ; Sequence 50, Application US/09898837A  
 ; Publication No. US2003007697A1  
 ; GENERAL INFORMATION:

APPLICANT: Quinn, Kerry E.  
 APPLICANT: Spytek, Kimberly A.  
 APPLICANT: Majumder, Kumud  
 APPLICANT: Vernet, Corine  
 APPLICANT: Herrmann, John L.  
 APPLICANT: Burgess, Catherine  
 APPLICANT: Fernandes, Elma  
 APPLICANT: Taupier Jr., Raymond  
 APPLICANT: Rastelli, Luca  
 APPLICANT: Curagen Corporation  
 APPLICANT: Gerlach, Valerie L.  
 APPLICANT: MacDougall, John R.  
 TITLE OF INVENTION: NOVEL SERINE/THREONINE PROTEIN-KINASE LIKE PROTEINS AND  
 FILE REFERENCE: 15966-598 CIP  
 CURRENT APPLICATION NUMBER: US/09/898,837A  
 PRIOR FILING DATE: 2001-07-03  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986  
 PRIOR FILING DATE: 1999-11-17  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839  
 PRIOR FILING DATE: 2000-04-05  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637  
 PRIOR FILING DATE: 2000-04-07  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080  
 PRIOR FILING DATE: 2000-04-13  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677  
 PRIOR FILING DATE: 2000-09-15  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347  
 PRIOR FILING DATE: 2000-02-09  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195  
 PRIOR FILING DATE: 2000-04-03  
 PRIOR APPLICATION NUMBER: U.S.S.N. 60/215,906  
 PRIOR FILING DATE: 2000-07-03  
 PRIOR APPLICATION NUMBER: U.S.S.N. 09/715,427  
 NUMBER OF SEQ ID NOS: 53  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 50  
 LENGTH: 215  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-898-837A-50

Alignment Scores: 6.68e-24 Length: 215  
 Pred. No.:

Score: 328.00 Matches: 79  
 Percent Similarity: 52.91% Conservative: 39  
 Best Local Similarity: 35.43% Mismatches: 93  
 Query Match: 21.03% Indels: 12  
 DB: 11 Gaps: 4

US-10-037-270-482 (1-866) x US-09-898-837A-50 (1-215)

```

Qy      69 CTGGCGGAGACATTTTCTTTCTGACATCATGTTGACAAAGAAAGAACCTGCTCCCTAT 128
Db      1 ILeValG1yGlyTrpThCysAlaAlaAsnSerIle-----ProTyr 14
Qy      129 TTGGTGTACTCTCAAGTCT---CACTTCAACCCCTGTGTGGGCGCTCTCATCAACCAAGC 185
Db      15 GlnValSerLeuAsnSerGlyIlePhe-----CysGlyGlySerLeuLeuAsnSerGln 32
Qy      186 TGGGTGTGGGCGGCGGCTGCTGCTATTTACCAAAATGCAAAAGTGAATGCTGGAAATTTG 245
Db      33 TrpValValSerAlaAlaIleCysTyrLysSerArgIleGlnValArgLeuGlyGluH18 52
Qy      246 AAGACAGAGTGAGAGGAGTACTGAGACAGCAATTAACCCGATTCAGATGCGCGCTAC 305
Db      53 AenIleAspValLeuGlnGlyAsnGlnGlnPheIleAsnAlaAlaLysIleIleThrH18 72
Qy      306 TGAAGTACAGTACATAGCGCCGCCACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 365
Db      73 ProAsnPheAsnGlyAsnThrLeuAsnAspIleMetLeuIleLysLeuSerSerPro 92
Qy      366 GCGATGCTCAATCCCAAGTCCAGCCCTTCCCTGCGCAGCACCAATGTCAGGCGAGC 425
Db      93 AlaThrLeuAsnSerArgValAlaThrValSerLeuProArgSerCysAlaAlaAlaGly 112
Qy      426 ACTGTCTGTCTACTCTGCTGTTGAGCTGAGCCCAAGAAAGAGTGGCCGACACCTGAC 485
Db      113 ThrGlnCysLeuIleSerGlyTrpGlyAsnTrpLysSerSerGlySerSerTrpProSer 132
Qy      486 TTGGCGGAGAACCTGAGAGGCGCCGCTGATGCTGATGCAAGTGCAGAAAGCAAGAACAA 545
Db      133 LeuLeuGlnCysLeuLysAlaProValLeuSerAspSerCysLysSerSerTrpPro 152
Qy      546 GGAAGAACGACAGAGAAATCTTATGTGTGAATTTGGAAGTATTCAGCCGAATTTT 605
Db      153 G1yGlnIleThrGlyAsnMetIleCysValGlyPheLeuGlnGlyLysAspSerCys 172
Qy      606 GGGAGAGTGGCGGCTGCTACTGATGCTGCAAGCAAGCAAGTCCAGGAGATC----- 656
Db      173 G1nglyAspSerGlyProValValCysAsnGlyGlnLeuGlnGlyIleValaSerTrp 192
Qy      657 GAGTGGGCGACTTCATGAGGAGGAGCGTGGCATCTACACCAATGTTTACAAATATGTA 716
Db      193 G1yTrpGlyCysAlaGlnLysAsnLysProGlyValaLysValaCysAsnTrpVal 212
Qy      717 TCCTGGATT 725
Db      213 AsnTrpIle 215

```

## RESULT 12

US-10-021-368-7  
 ; Sequence 7, Application US/10021368  
 ; Publication No. US20020106367A1  
 ; GENERAL INFORMATION:

APPLICANT: Band, Wimla  
 TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
 MOLECULES AND METHODS

NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:



```

Oy      423 GGCACTGTCGTACTTACATCGATTGGAGCCAGAAGAAAAGTGCGCACACCCT 482
         ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      113 GLythrInGcYsbLeuIleSerGlyTrpGLysAenThrLylSereSerGlyThrSerTyrrPro 132
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      483 GACTTGCAGCAGAACCTGGAGGCCCCCCTGATGTCTCATGTGAATATTCAGAAATCCAAAAACAGAA 542
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      153 ProGIgYLlneHrShserAmnecPheCySaLaGlYtRleuIdUjglYlysAepSer 172
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      603 TTTGGGAGAGTGCCCGTTGTCTACTGTGCATCTGCAACAAGACAGCTCCAGGAAATC----- 656
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      173 CysGIngLYAspsSeGrCYglYPrOVAValCYseScdYlYlsbeutIngLYllevalSer 192
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      657 --GAGTGGGGCACCTTCATGGGAGGGGACGTGCGCATCTACCAAATGTTTAACAATAAT 713
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      193 TrpGIseRGlyCYsaLaGlNlySAeNuYSProGlyVAITyrThrySVAlCYSAenTYr 212
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      714 GTATCCTGGATGTGAGAACAACCT 734
         ||||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      213 ValSerTpILelEyslgnThr 219
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 14
US-10-239-663-39
; Sequence 39, Application US/10239663
; Publication NO. US20030139572A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia, K.
; APPLICANT: Smith, Randall, F.
; APPLICANT: Xiang, Zhaoying
; APPLICANT: Kahnlick, Karen
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GPS0018
; CURRENT APPLICATION NUMBER: US/10/239,663
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/192,158
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,668
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/200,166
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-239-663-39

Alignment Scores:
Pred. No.:          3,456-22           Length:       243
Score:             311.00              Matches:      85
Percent Similarity: 50.81%              Conserves:     41
Best Local Similarity: 34.27%            Mismatches:    104
Query Match:       19.94%               Indels:       18
DB:                12                   Gaps:         7

US-10-037-270-482 (1-866) x US-10-239-663-39 (1-243)

Oy      27 AAGACCTGCATCCATCAATAATATGCTCTTAATTTGGGATGTCCTGCCTGGACATTTC 86
         ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      4   LysSetrrApSnPheluSerMetLeuLeuPheProVaIalaleuXla-----Phe 20
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Oy      87 TTTCGTGACTCATCTGTTTCAGAAAGAACCCCTGCCTCCCATTTTGGTAGTCACTCAAGTCT 146
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      21 AenPrObepTYrThrvAl-----SerSerThrProPriotyZreuvalltyrLeuVSSer 38
         :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

[illegible]

; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 248  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-137-870-248

Alignment Scores:  
Pred. No.: 1.37e-21 Length: 241  
Score: 305.00 Matches: 83  
Percent Similarity: 50.41% Conservative: 41  
Best Local Similarity: 33.74% Mismatches: 104  
Query Match: 19.55% Indels: 18  
DB: 12 Gaps: 7

US-10-037-270-482 (1-866) x US-10-137-870-248 (1-241)

```

QY 42 ATGAAATATGCTTC-----TATTGGGTCCTCGCTGGACATTTTCTTGTCT 92
DB 1 MetLysPheIleuLeuTPAlaLeuLeuAsnLeuThrValAlaLeuAlaPheAsnPro 20
QY 93 GACTCATCTGTGAGAAAGAACCCCTGCTCCCTATTGGTGTACTCAAGTCTCACTTC 152
DB 21 AspTyrThrVal-----SerSerThrProProTyrLeuValTyrLeuLysSerAspTyr 38
QY 153 AACCCCTGTGTGGGCGTCTCATCAAAACCCAGCTGGTGTGGCCGCTCACTGCTAT 212
DB 39 LeuProCysAlaGlyValLeuIleHisProLeuTyrValIleThrAlaAlaHisCysAsn 58
QY 213 TTACCAATCTGAAAGATGAGTGGGAAATTTCAGAGCAGAGTCAGAGCGGTACTGAA 272
DB 59 LeuProLysLeuArgValIleLeuGly-----ValThrIleProAlaAspSerAsnGlu 76
QY 273 -----CAGCAATTAACCCCATTCAGATGCTCGCTACTGAGACTACATACATAGC 323
DB 77 LysHisLeuGlnValIleGlyTyrGlnLysMetIleHisProHisPheSerValThr 96
QY 324 GCCCCACAGATGACTCATGCTCATCAAGCTGAGCTTAAGCTGCTCAATCCCAA 383
DB 97 SerIleAspHisAspIleMetLeuIleLysLeuLysThrGlnAlaGlnLeuAsnAspTyr 116
QY 384 GTCCAGCCCTTCCCTCGCCACCAACAATGTCAAGCCAGGCACTGTGTCTACTCTCA 443
DB 117 ValLysLeuAlaAsnLeuProTyrGlnThrIleSerGlnAsnThrMetCysSerValSer 136
QY 444 GATTGGACTGAGCCAGAAACAGTGGCCGACACCTGACTTGGCGGAGAACTGTGAG 503
DB 137 ThrTTrpSerTyrAsnValCysAspIleTyrLysGlnProAspSerLeuGlnThrValAsn 156
QY 504 GCCCCCGTATGTCGTATCGAATGCCAATAAAACAGAACAGAAAGCCAC----- 557
DB 157 IleSerValIleSerLysProGlnCysArg-----AspAlaTyrLysThrTyrAsnIle 174
QY 558 ---AGGAATTCCTTATGTGAAATTTGAAAGATATTCAACCGAATTTTGGGGAGGTG 614
DB 175 ThrGlnAsnMetLeuCysValGlyIleValIleProGlyArgArgGlnProCysLysGlnVal 194
QY 615 GCGGTGCTACTCTCATCTGCAAAAGACAGCTCCAGGGAATC-----GAGGTGGGG 665
DB 195 SerAlaIleProAlaIleCysAsnGlyMetLeuGlnGlyIleLeuSerPheAlaAspGly 214
QY 666 CACTTCATGGAGGAGGAGTGGGATCTTACCAATGATTTACAAATATGATCTGATTT 725
DB 215 CysValLeuArgAlaAspValGlyIleTyrAlaLysIlePheTyrTyrIleProTyrIle 234
QY 726 GAGAAACATGCTAAGAC 743
DB 235 GluAsnValIleGlnAsn 240

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: November 7, 2003, 23:05:46 ; Search time 3495 Seconds  
(without alignments)  
10136.691 Million cell updates/sec

Title: US-10-037-270-482  
Perfect score: 866  
Sequence: 1 ggcacttactccctgagctca.....tcgggaatgtagcactactagt 866

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813366 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_bai:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_hcg\_hum:\*  
39: em\_hcg\_mus:\*  
40: em\_hcg\_other:\*  
41: em\_hcg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862.8	99.6	1060	AX455324	AX455324 Sequence
2	857.8	99.1	890	AX424630	AX424630 Sequence
3	850.8	98.2	865	AR240728	AR240728 Sequence
4	850.8	98.2	865	AX318279	AX318279 Sequence
5	834	96.3	865	AR263885	AR263885 Sequence
6	730	84.3	730	AX350386	AX350386 Sequence
7	728.4	84.1	730	AX350388	AX350388 Sequence
8	706.4	81.6	708	AX247864	AX247864 Sequence
9	706.4	81.6	708	AX455326	AX455326 Sequence
10	703.8	81.3	721	AX350481	AX350481 Sequence
11	606.4	70.0	798	AX360087	AX360087 Sequence
12	527	60.9	1855	AK097648	AK097648 Sequence
13	511.2	59.0	705	AR240729	AR240729 Sequence
14	511.2	59.0	705	AX318281	AX318281 Sequence
15	359.8	41.5	479	BD113886	BD113886 EST and e
16	259	29.9	31047	AC093638	AC093638 Homo sapi
17	246	28.4	379	BD047802	BD047802 Sequence
18	196.8	22.7	136812	AC137710	AC137710 Mus muscu
19	196.8	22.7	204476	AC117613	AC117613 Mus muscu
20	193	22.3	225362	AC094695	AC094695 Rattus no
21	173.6	20.0	804	RNTRY1	V01273 Rat mRNA en
22	169.4	19.6	814	NMTRYAR	X04574 Mouse mRNA
23	167.8	19.4	860	GGU15156	U15156 Gallus gall
24	167.8	19.4	864	GGU15155	U15155 Gallus gall
25	164.4	19.0	797	AB009661	AB009661 Mus muscu
26	159.6	18.4	773	RNTRY2	V01274 Rat mRNA en
27	155.4	17.9	805	BTTRYAP	X54703 Bovine mRNA
28	154.4	17.8	741	HMUMA	D45417 Human mRNA
29	154.4	17.8	790	E15808	E15808 Human mRNA
30	154.4	17.8	821	BC030238	BC030238 Homo sapi
31	154.4	17.8	853	HSTRP1V	X72781 Homo sapien
32	154.4	17.8	875	BC030260	BC030260 Homo sapi
33	154	17.8	850	AX333266	AX333266 Sequence
34	154	17.8	850	HSTRYIVB	X71345 H.sapiens m
35	153.2	17.7	802	AX527621	AX527621 Sequence
36	153.2	17.7	802	HUMTRPSGNB	M27602 Human pancr
37	153.2	17.7	954	AX014870	AX014870 Sequence
38	152.8	17.6	741	E01617	E01617 cDNA encodi
39	152.8	17.6	744	E09633	E09633 DNA encodin
40	152.8	17.6	807	HSTRYIII	X15505 Human mRNA
41	145	16.7	825	BOVPCP	D38507 Bos taurus
42	145	16.7	825	AR103928	AR103928 Sequence
43	144.6	16.7	819	DOGTRYPA	M11589 Dog pancrea
44	142.8	16.5	860	GGU15157	U15157 Gallus gall
45	142	16.4	699	AX206893	AX206893 Sequence

## ALIGNMENTS

RESULT 1  
AX455324  
LOCUS AX455324 1060 bp DNA linear PART 06-JUL-2002  
DEFINITION Sequence 1 from Patent WO0206331.  
ACCESSION AX455324  
VERSION AX455324.1 GI:21714507  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Meyers, R.  
TITLE 14087, a novel serine protease molecule and uses therefor  
JOURNAL Patent: WO 0206331-A 1 24-JAN-2002;

FEATURES  
source  
Millennium Pharmaceuticals, Inc. (US)  
Location/Qualifiers  
1.1060  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
164.871  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD38234.1"  
/db\_xref="GI:21714508"  
translation="MKTVFYGLVLAAGTFFPVDSSVQEDPAPLVYIKSHFNPVGL  
IKPSWLAAPHCYLPNLKWLGNPKSRVRDQETNPIDQIVATMYSHSAPQDDML  
IKLAKPAMLPKVPPLPLATNVRPGVCLSLGSDMSQENSGHPDLRLQLEAPVMS  
RECOETEOGKSHNSLCKVFKFVSRIFEVAVATYCKDKLGIIEVGHMGSDVGIY  
TNYKYVSMENIAKDK"

BASE COUNT 309 a 259 c 237 g 251 t 4 others

ORIGIN

Query Match 99.8%; Score 862.8; DB 6; Length 1060;  
Best Local Similarity 99.8%; Pred. No. 2,7e-232;  
Matches 864; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGCACTTACTCTCCCTGAGCTAAGGGGGAAGAGCTGATCCACATGAATATGCTTATT 60  
DB 123 GACACTTACTCTCCCTGAGCTAAGGGGGAAGAGCTGATCCACATGAATATGCTTATT 182  
QY 61 TGGGTGCTCCGCTGGGGAATTTTCTTGTGCTGATCTGTCAGAAAGAACCCCTG 120  
DB 183 TGGGTGCTCCGCTGGGGAATTTTCTTGTGCTGATCTGTCAGAAAGAACCCCTG 242  
QY 121 CTCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 180  
DB 243 CTCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 302  
QY 181 CCAGCTGGGTGCTGGGCCCCAGCTCACTGCTATTACCAATCTGAAGTATGCTGGAA 240  
DB 303 CCAGCTGGGTGCTGGGCCCCAGCTCACTGCTATTACCAATCTGAAGTATGCTGGAA 362  
QY 241 ATTTCAAGAGCAAGTCAAGAGCGTATGTAACAAGCAATTAACCCATCAATGCTCC 300  
DB 363 ATTTCAAGAGCAAGTCAAGAGCGTATGTAACAAGCAATTAACCCATCAATGCTCC 422  
QY 301 GCTACTGSACTACAGTCAATAGCGCCCAAGAGTGAATGCTATCAAGTGGCTA 360  
DB 423 GCTACTGSACTACAGTCAATAGCGCCCAAGAGTGAATGCTATCAAGTGGCTA 482  
QY 361 AGCTGCGATGCTCAATCCCAAGTCAAGCCCTTCCCTGCGCACCAATGTCAGGC 420  
DB 483 AGCTGCGATGCTCAATCCCAAGTCAAGCCCTTCCCTGCGCACCAATGTCAGGC 542  
QY 421 CAGGCACTGCTGTTCACTCACTGTTGTAAGTGAAGCAAGAAACAGTGGCCGAC 480  
DB 543 CAGGCACTGCTGTTCACTCACTGTTGTAAGTGAAGCAAGAAACAGTGGCCGAC 602  
QY 481 CTGACTTGGGCGAGAACTGGAAGGCCCGCTGATGTTGATCGAAGATGCAAAAAAG 540  
DB 603 CTGACTTGGGCGAGAACTGGAAGGCCCGCTGATGTTGATCGAAGATGCAAAAAAG 662  
QY 541 AACAGAGAAAAGCCCAAGGAATTCCTTATGTTGAAATTTGTAAGTATGACCGAA 600  
DB 663 AACAGAGAAAAGCCCAAGGAATTCCTTATGTTGAAATTTGTAAGTATGACCGAA 722  
QY 601 TTTTGGGAGAGTGGCGCTGTTGCTACTGTCATCTGCAAGAACAGCTCAGGGAATGAGG 660  
DB 723 TTTTGGGAGAGTGGCGCTGTTGCTACTGTCATCTGCAAGAACAGCTCAGGGAATGAGG 782  
QY 661 TGGGGCACTTCAATGGAAGGGGAGTGGGCACTTACCAATGTTTCAATATGATTCCT 720  
DB 783 TGGGGCACTTCAATGGAAGGGGAGTGGGCACTTACCAATGTTTCAATATGATTCCT 842  
QY 721 GGAATTGAGAACCTGCTAAGAGCAAGTGAAGCCCTATCTCCCTGCTGATTCACATGCG 780

DB 843 GGAATTGAGAACCTGCTAAGAGCAAGTGAAGCCCTATCTCTCCCTGCTGATTCACCTGGC 902  
QY 781 TCTGCACTGAGACTATCAAGCAGATATTTTCCCTGATTCATTAATTAATTCGAATG 840  
DB 903 TCTGCACTGAGACTATCAAGCAGATATTTTCCCTGATTCATTAATTAATTCGAATG 962  
QY 841 AAAATTTGGGAATGTAGCATCTAGT 866  
DB 963 AAAATTTGGGAATGTAGCATCTAGT 988

RESULT 2  
AX342630 890 bp DNA linear PAT 12-JAN-2002  
LOCUS Sequence 27 from Patent WO0198468.  
DEFINITION AX342630  
ACCESSION AX342630  
VERSION AX342630.1 GI:18152027  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
1 Yue, H., Elliott, V. S., Gandhi, A. R., Lal, P., Au-Young, J.,  
Tribouley, C. M., Deleane, A. M., Baughn, M. R., Nguyen, D. B., Lee, E. A.,  
Hafalia, A., Khan, F. A., Wallis, N. K., Yao, M. G., Lu, D. A., Patterson, C.,  
Tang, Y. T., Walsh, R. T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R.  
Patent: WO 0198468-A 27-DEC-2001;  
Incyte Genomics, Inc. (US)

FEATURES  
source  
1.890  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7474081B1"

BASE COUNT 249 a 230 c 203 g 208 t

ORIGIN

Query Match 99.1%; Score 857.8; DB 6; Length 890;  
Best Local Similarity 99.8%; Pred. No. 6.8e-231;  
Matches 859; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGCACTTACTCTCCCTGAGCTAAGGGGGAAGAGCTGATCCACATGAATATGCTTATT 60  
DB 22 GGCACTTACTCTCCCTGAGCTAAGGGGGAAGAGCTGATCCACATGAATATGCTTATT 81  
QY 61 TGGGTGCTCCGCTGGGGAATTTTCTTGTGCTGATCTGTCAGAAAGAACCCCTG 120  
DB 82 TGGGTGCTCCGCTGGGGAATTTTCTTGTGCTGATCTGTCAGAAAGAACCCCTG 141  
QY 121 CTCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 180  
DB 142 CTCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 201  
QY 181 CCAGCTGGGTGCTGGGCCCCAGCTCACTGCTATTACCAATCTGAAGTGAATGCTGGAA 240  
DB 202 CCAGCTGGGTGCTGGGCCCCAGCTCACTGCTATTACCAATCTGAAGTGAATGCTGGAA 261  
QY 241 ATTTCAAGAGCAAGTCAAGAGCGTATGTAACAAGCAATTAACCCATTCAGATGCTCC 300  
DB 262 ATTTCAAGAGCAAGTCAAGAGCGTATGTAACAAGCAATTAACCCATTCAGATGCTCC 321  
QY 301 GCTACTGSACTACAGTCAATAGCGCCCAAGAGTGAATGCTATGCTCAAGTGGCTA 360  
DB 322 GCTACTGSACTACAGTCAATAGCGCCCAAGAGTGAATGCTATGCTCAAGTGGCTA 381  
QY 361 AGCTGCGATGCTCAATCCCAAGTCAAGCCCTTCCCTGCGCACCAATGTCAGGC 420  
DB 382 AGCTGCGATGCTCAATCCCAAGTCAAGCCCTTCCCTGCGCACCAATGTCAGGC 441  
QY 421 CAGGCACTGCTGTTCACTCACTGTTGTAAGTGAAGCAAGAAACAGTGGCCGAC 480  
DB 442 CAGGCACTGCTGTTCACTCACTGTTGTAAGTGAAGCAAGAAACAGTGGCCGAC 501





BASE COUNT 235 a 226 c 196 g 208 t  
ORIGIN  
Query Match 98.2%; Score 850.8; DB 6; Length 865;  
Best Local Similarity 99.7%; Pred. No. 6.5e-229;  
Matches 863; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 1 GGCACTTACTCCCTGAGCTAAGGGGGAAGAGCTGATCAACATGAATATGTCTTATT 60  
DB 1 GGCACTTACTCCCTGAGCTAAGGGGGAAGAGCTGATCAACATGAATATGTCTTATT 60  
QY 61 TGGGTCTCTGCTGGGACATTTTCTTGTCTGATCTCATCTGTTGAGAAAGAACCTTG 120  
DB 61 TGGGTCTCTGCTGGGACATTTTCTTGTCTGATCTCATCTGTTGAGAAAGAACCTTG 120  
QY 121 CTCCTATTGTTGGTGAACCTCAAGCTCACTTCAACCCCTGTGGGGCTCTCATCAAC 180  
DB 121 CTCCTATTGTTGGTGAACCTCAAGCTCACTTCAACCCCTGTGGGGCTCTCATCAAC 180  
QY 181 CCAGCTGGGTGCTGGCCCGAGCTCACTGCTATTACCAATCTGAAGATGATGCTGGAA 240  
DB 181 CCAGCTGGGTGCTGGCCCGAGCTCACTGCTATTACCAATCTGAAGATGATGCTGGAA 240  
QY 241 ATTTCAAGAGCAGAGTCAAGACGGTACTGACAGACAAATTAACCCCATTCAGATGCTC 300  
DB 241 ATTTCAAGAGCAGAGTCAAGACGGTACTGACAGACAAATTAACCCCATTCAGATGCTC 300  
QY 301 GCTACTGGAACCTACAGTCAAGAGGCCCCCAAGATGATCTCATCTGATCAAGCTGCTA 360  
DB 301 GCTACTGGAACCTACAGTCAAGAGGCCCCCAAGATGATCTCATCTGATCAAGCTGCTA 360  
QY 361 AGCTCCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGGCCAGCCCAATGTCAGG 420  
DB 361 AGCTCCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGGCCAGCCCAATGTCAGG 420  
QY 421 CAGGCACTGCTCTGCTACTCTCAAGTTTGAAGCTGAGCCAAAGAACAGTGGCCGAC 480  
DB 421 CAGGCACTGCTCTGCTACTCTCAAGTTTGAAGCTGAGCCAAAGAACAGTGGCCGAC 480  
QY 481 CTGACTTGGCGGAGAACTGGAAGGCCCCCGTGAATGCTGATGAGAAATGCAAAAAAC 540  
DB 481 CTGACTTGGCGGAGAACTGGAAGGCCCCCGTGAATGCTGATGAGAAATGCAAAAAAC 540  
QY 541 AACAAAGAAAAAGCCACAGGAATTCCTTATGTGTGAATTTGTGAAGTATTCAGCCGA 600  
DB 541 AACAAAGAAAAAGCCACAGGAATTCCTTATGTGTGAATTTGTGAAGTATTCAGCCGA 600  
QY 601 TTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGAACAGCTCCAGGAATCGAG 660  
DB 601 TTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGAACAGCTCCAGGAATCGAG 660  
QY 661 TGGGGCACTTCATGGAAGGGAAGCTGGGCACTTACACCAATGTTTACAAATATGATCT 720  
DB 661 TGGGGCACTTCATGGAAGGGAAGCTGGGCACTTACACCAATGTTTACAAATATGATCT 720  
QY 721 GGAATGAGAAACCTGCTAAGGACAAGTGAACCTTCTCCCTGTCATTCAGCTGGC 780  
DB 721 GGAATGAGAAACCTGCTAAGGACAAGTGAACCTTCTCCCTGTCATTCAGCTGGC 780  
QY 781 TCTGCAATGACTATACAGCAGATATATTTTCCCTCTATTCAAATTAATTCCTCAAT 840  
DB 781 TCTGCAATGACTATACAGCAGATATATTTTCCCTCTATTCAAATTAATTCCTCAAT 840  
QY 841 AAAATTTGGGAATGTAGCAATCTAGT 866  
DB 841 AAAATTTGGGAATGTAGCAATCTAGT 866

RESULT 5  
AR263885/c AR263885 865 bp DNA linear PAT 29-JAN-2003  
LOCUS AR263885  
DEFINITION Sequence 63 from patent US 6331427.  
ACCESSION AR263885

VERSION AR263885.1 GI:28075889  
KEYWORDS  
SOURCE unknown.  
ORGANISM unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 865)  
TITLES Robison, K.E.  
JOURNAL Protease homologs  
FEATURES Patent: US 6331427-A 63 18-DEC-2001;  
source Location/Qualifiers  
1..865  
/organism="unknown"  
BASE COUNT 207 a 197 c 225 g 236 t  
ORIGIN  
Query Match 96.3%; Score 834; DB 6; Length 865;  
Best Local Similarity 98.7%; Pred. No. 3.6e-224;  
Matches 851; Conservative 0; Mismatches 10; Indels 1; Gaps 1;  
QY 1 GGCACTTACTCCCTGAGCTAAGGGGGAAGAGCTGATCAACATGAATATGTCTTATT 60  
DB 865 GGCACTTACTCCCTGAGCTAAGGGGGAAGAGCTGATCAACATGAATATGTCTTATT 866  
QY 61 TGGGTCTCTGCTGGGACATTTTCTTGTCTGATCTCATCTGTTGAGAAAGAACCTTG 120  
DB 805 TGGGTCTCTGCTGGGACATTTTCTTGTCTGATCTCATCTGTTGAGAAAGAACCTTG 120  
QY 121 CTCCTATTGTTGGTGAACCTCAAGCTCACTTCAACCCCTGTGGGGCTCTCATCAAC 180  
DB 745 CTCCTATTGTTGGTGAACCTCAAGCTCACTTCAACCCCTGTGGGGCTCTCATCAAC 180  
QY 181 CCAGCTGGGTGCTGGCCCGAGCTCACTGCTATTACCAATCTGAAGATGATGCTGGAA 240  
DB 685 CCAGCTGGGTGCTGGCCCGAGCTCACTGCTATTACCAATCTGAAGATGATGCTGGAA 240  
QY 241 ATTTCAAGAGCAGAGTCAAGACGGTACTGACAGACAAATTAACCCCATTCAGATGCTC 300  
DB 625 ATTTCAAGAGCAGAGTCAAGACGGTACTGACAGACAAATTAACCCCATTCAGATGCTC 300  
QY 301 GCTACTGGAACCTACAGTCAAGAGGCCCCCAAGATGATCTCATCTGATCAAGCTGCTA 360  
DB 565 GCTACTGGAACCTACAGTCAAGAGGCCCCCAAGATGATCTCATCTGATCAAGCTGCTA 360  
QY 361 AGCTCCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGGCCAGCCCAATGTCAGG 419  
DB 505 AGCTCCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGGCCAGCCCAATGTCAGG 419  
QY 420 CCAGGCACTGCTCTGCTACTCTCAAGTTTGAAGCTGAGCCAAAGAACAGTGGCCGAC 479  
DB 445 CCAGGCACTGCTCTGCTACTCTCAAGTTTGAAGCTGAGCCAAAGAACAGTGGCCGAC 479  
QY 480 CCGTACTTGGCGGAGAACTGGAAGGCCCCCGTGAATGCTGATGAGAAATGCAAAAAAC 539  
DB 385 CCGTACTTGGCGGAGAACTGGAAGGCCCCCGTGAATGCTGATGAGAAATGCAAAAAAC 539  
QY 540 GAAACAAGAAAAAGCCACAGGAATTCCTTATGTGTGAATTTGTGAAGTATTCAGCCGA 599  
DB 325 GAAACAAGAAAAAGCCACAGGAATTCCTTATGTGTGAATTTGTGAAGTATTCAGCCGA 599  
QY 600 ATTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGAACAGCTCCAGGAATCGAG 659  
DB 265 ATTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGAACAGCTCCAGGAATCGAG 659  
QY 660 GTGGGGCACTTCATGGAAGGGAAGCTGGGCACTTACACCAATGTTTACAAATATGATCT 719  
DB 205 GTGGGGCACTTCATGGAAGGGAAGCTGGGCACTTACACCAATGTTTACAAATATGATCT 719  
QY 720 TGGATGAGAAACCTGCTAAGGACAAGTGAACCCCTCTCTCCCTGTCATTCAGCTGG 779  
DB 145 TGGATGAGAAACCTGCTAAGGACAAGTGAACCCCTCTCTCCCTGTCATTCAGCTGG 779  
QY 780 CTCCTGCAATGACTATACAGCAGATATATTTTCCCTCTATTCAAATTAATTCCTCAAT 839  
DB 780 CTCCTGCAATGACTATACAGCAGATATATTTTCCCTCTATTCAAATTAATTCCTCAAT 839

Db 85 CTCTGCCATGAGATATACAGACGATATTTCTCTATTCAATAAATTCCTCAAT 26  
Qy 840 GAAATTTGGGAATGATGACATA 861  
Db 25 GAAATTTGGGAATGATGACAA 4

RESULT 6  
AX350386  
LOCUS AX350386 730 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 11 from Patent WO0181578.  
ACCESSION AX350386  
VERSION AX350386.1 GI:18616037  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Verne, C.A., Fernandes, E.R., Gerlach, V., Shinkets, R.A.,  
Malpanker, U.M., Boldog, F.L., Zerkusen, B.D., Spyrek, K.A.,  
Majumder, K., Tchernev, V.T., Padigar, M., Patcurajan, M.,  
Burgess, C.E., Gangoli, E.A., Smithson, G., Rastelli, L.,  
MacDougall, J.R., Taupier, R.J., Grose, W.M., and Alsobrook, J.P.  
TITLE Novel proteins and nucleic acids encoding same  
JOURNAL Patent: WO 0181578-A 11 01-NOV-2001;  
Curegen Corporation (US)  
FEATURES  
source Location/Qualifiers  
1..730  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 191 a 197 c 170 g 172 t  
ORIGIN

Query Match 84.3%; Score 730; DB 6; Length 730;  
Best Local Similarity 100.0%; Pred. No. 8.5e-195;  
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 GATCACCATGAATATATGTCTTCTATTGGGTGCTCGCTGGACATTTTCTTGCTGA 94  
Db 1 GATCACCATGAATATATGTCTTCTATTGGGTGCTCGCTGGACATTTTCTTGCTGA 60

Qy 95 CTCATCTGTTCAAGAAAGAACCCCTGCTCCCTATTGGTGTACTCAAGCTCACTTCA 154  
Db 61 CTCATCTGTTCAAGAAAGAACCCCTGCTCCCTATTGGTGTACTCAAGCTCACTTCA 120

Qy 155 CCCCTGTGTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCACTCATGCTATT 214  
Db 121 CCCCTGTGTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCACTCATGCTATT 180

Qy 215 ACCAATCTGAAGATGATGCTGGAAATTTCAAGACAGATCAGAGCGGTACTGAAC 274  
Db 181 ACCAATCTGAAGATGATGCTGGAAATTTCAAGACAGATCAGAGCGGTACTGAAC 240

Qy 275 GACAAATTAACCCATTCAGATGCTCCGCTAAGTGAATCAAGTCAAGGCCCAAGGA 334  
Db 241 GACAAATTAACCCATTCAGATGCTCCGCTAAGTGAATCAAGTCAAGGCCCAAGGA 300

Qy 335 TGAACCTCATGCTCATCAAGCTGCTAAGCTGCTCATGCTCAATCCCAAGTCCAGCCCT 394  
Db 301 TGAACCTCATGCTCATCAAGCTGCTAAGCTGCTCATGCTCAATCCCAAGTCCAGCCCT 360

Qy 395 TCCCTGCGCACCAACCAATGCTCAGGCCAGGACCTGCTGTCTAATCTCAAGTTGAGCTG 454  
Db 361 TCCCTGCGCACCAACCAATGCTCAGGCCAGGACCTGCTGTCTAATCTCAAGTTGAGCTG 420

Qy 455 GAGCCAAAGAAAGAGTGGCGAGACCCCTGACTGGCGAGAACTGGAAGGCCCGCTGAT 514  
Db 421 GAGCCAAAGAAAGAGTGGCGAGACCCCTGACTGGCGAGAACTGGAAGGCCCGCTGAT 480

Qy 515 GTCTGATCGAGATGCGCAAAAACAGAACAGAAAGAAAGCAGAGAAATTCCTTATGTGT 574

Db 481 GTCTGATCGAGATGCGCAAAAACAGAACAGAAAGAAAGCAGAGAAATTCCTTATGTGT 540  
Qy 575 GAAATTTGGAAGATATTGAGCGGATTTTGGGGAGGTGGCGGTGCTACTGCTATG 634  
Db 541 GAAATTTGGAAGATATTGAGCGGATTTTGGGGAGGTGGCGGTGCTACTGCTATG 600

Qy 635 CAAAGCAAGCTCCAGGAATGAGGTGGGACCTTCATGAGAGGGAGCGTGGCATCTA 694  
Db 601 CAAAGCAAGCTCCAGGAATGAGGTGGGACCTTCATGAGAGGGAGCGTGGCATCTA 660

Qy 695 CACCAATGTTTCAAAATATGATCTCGATTTGAGAACACTGCTTAAGACAAAGTGAACCC 754  
Db 661 CACCAATGTTTCAAAATATGATCTCGATTTGAGAACACTGCTTAAGACAAAGTGAACCC 720

Qy 755 TACTTCTCCC 764  
Db 721 TACTTCTCCC 730

RESULT 7  
AX350388  
LOCUS AX350388 730 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 13 from Patent WO0181578.  
ACCESSION AX350388  
VERSION AX350388.1 GI:18616038  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Verne, C.A., Fernandes, E.R., Gerlach, V., Shinkets, R.A.,  
Malpanker, U.M., Boldog, F.L., Zerkusen, B.D., Spyrek, K.A.,  
Majumder, K., Tchernev, V.T., Padigar, M., Patcurajan, M.,  
Burgess, C.E., Gangoli, E.A., Smithson, G., Rastelli, L.,  
MacDougall, J.R., Taupier, R.J., Grose, W.M., and Alsobrook, J.P.  
TITLE Novel proteins and nucleic acids encoding same  
JOURNAL Patent: WO 0181578-A 13 01-NOV-2001;  
Curegen Corporation (US)  
FEATURES  
source Location/Qualifiers  
1..730  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 191 a 198 c 170 g 171 t  
ORIGIN

Query Match 84.1%; Score 728.4; DB 6; Length 730;  
Best Local Similarity 99.9%; Pred. No. 2.4e-194;  
Matches 729; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 35 GATCACCATGAATATATGTCTTCTATTGGGTGCTCGCTGGACATTTTCTTGCTGA 94  
Db 1 GATCACCATGAATATATGTCTTCTATTGGGTGCTCGCTGGACATTTTCTTGCTGA 60

Qy 95 CTCATCTGTTCAAGAAAGAACCCCTGCTCCCTATTGGTGTACTCAAGCTCACTTCA 154  
Db 61 CTCATCTGTTCAAGAAAGAACCCCTGCTCCCTATTGGTGTACTCAAGCTCACTTCA 120

Qy 155 CCCCTGTGTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCACTCATGCTATT 214  
Db 121 CCCCTGTGTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCACTCATGCTATT 180

Qy 215 ACCAATCTGAAGATGATGCTGGAAATTTCAAGACAGATCAGAGCGGTACTGAAC 274  
Db 181 ACCAATCTGAAGATGATGCTGGAAATTTCAAGACAGATCAGAGCGGTACTGAAC 240

Qy 275 GACAAATTAACCCATTCAGATGCTCCGCTAAGTGAATCAAGTCAAGGCCCAAGGA 334  
Db 241 GACAAATTAACCCATTCAGATGCTCCGCTAAGTGAATCAAGTCAAGGCCCAAGGA 300

Qy 335 TGAACCTCATGCTCATCAAGCTGCTAAGCTGCTCAATCCCAAGTCCAGCCCT 394

Db 301 TGACCTATGCTATCAAGCTGCGTAAGCTGCGCATCTCAATCCAAAGTCAGCCCT 360  
QY 395 TCCCTGCGCAGCAACCAATGTCAGGCGCAGCACTGCTGTCTACTCTCAGATTGACTG 454  
Db 361 TCCCTGCGCAGCAACCAATGTCAGGCGCAGCACTGCTGTCTACTCTCAGATTGACTG 420  
QY 455 GAGCCAGAAAACAGTGGCGGACACCTGACTTGGCGGAGAACTGGAGGCCCGCTGAT 514  
Db 421 GAGCCAGAAAACAGTGGCGGACACCTGACTTGGCGGAGAACTGGAGGCCCGCTGAT 480  
QY 515 GTCTGATCGAAGATGCGCAAAAACAGAAAGCAAGCAAGCAAGTCTTATGCT 574  
Db 481 GTCTGATCGAAGATGCGCAAAAACAGAAAGCAAGCAAGTCTTATGCT 540  
QY 575 GAAATTTGTAAAGATATTCAGCCGAATTTTGGGAGAGTGCCGCTTCTACTGCTCATCTG 634  
Db 541 GAAATTTGTAAAGATATTCAGCCGAATTTTGGGAGAGTGCCGCTTCTACTGCTCATCTG 600  
QY 635 CAAAGACACTCCAGGGAATCGAGTGGGCACTTATGAGAGGGAAGCTCGGCACTTA 694  
Db 601 CAAAGACACTCCAGGGAATCGAGTGGGCACTTATGAGAGGGAAGCTCGGCACTTA 660  
QY 695 CACCAATGTTTACAAATATGATTCCTGATTTGAGAACACTGTAAGCAAGTGAAGCC 754  
Db 661 CACCAATGTTTACAAATATGATTCCTGATTTGAGAACACTGTAAGCAAGTGAAGCC 720  
QY 755 TACTCTGCC 764  
Db 721 TACTCTGCC 730

RESULT 8  
AX247864 708 bp DNA linear PAT 28-SEP-2001  
LOCUS Sequence 105 from Patent WO0166748.  
DEFINITION AX247864  
ACCESSION AX247864 GI:15862497  
VERSION AX247864.1 GI:15862497  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukayoka; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 Conklin,D.C., Presnell,S.R. and Adler,D.A.  
AUTHORS Full length expressed human polynucleotides and the polypeptides  
TITLE they encode  
JOURNAL Patent: WO 0166748-A,105 13-SEP-2001;  
ZymoGenetics, Inc. (US)

FEATURES  
source Location/Qualifiers  
1..708  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
1..708  
/note="unnamed protein product"

CDS  
/codon\_start=1  
/protein\_id="CAC8648.1"  
/db\_xref="GI:15862498"

translation="MKYVPTLSVLAGTFPPADSSVQKEDPAPLYLVKHPNCGVL  
IKSWLAPAHCTLPNLKWLGNPKSRVRGTSQTINPIQIVYMYSHAPDDML  
IKLAKPAMLPKQPLTLATTVNRPGIVCLSLDQSEKSHPLRLQLEAPVNS  
RECOKTEOGKSHNSLCKVEKVFRIFGVAVALICCKDKLOGIEVGHFMGDVGIY  
TNYKYVSWIENIAXDK"

BASE COUNT 188 a 185 c 168 g 167 t  
ORIGIN  
Query Match 81.6%; Score 706.4; DB 6; Length 708;  
Best Local Similarity 99.9%; Pred. No. 3.9e-188;  
Matches 707; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 42 AAGAAATATGCTCTTATTTGGGTGCTCTGCGTGGACATTTTCTTGTGACTCATCT 101  
Db 1 AAGAAATATGCTCTTATTTGGGTGCTCTGCGTGGACATTTTCTTGTGACTCATCT 60

QY 102 GTTCAGAAAAGAACCCCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT 161  
Db 61 GTTCAGAAAAGAACCCCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT 120  
QY 162 GTGGGGCTCTCATCAAAACCCAGCTGGGTGCTGGCCCAAGCTACTGCTATTTACCAAT 221  
Db 121 GTGGGGCTCTCATCAAAACCCAGCTGGGTGCTGGCCCAAGCTACTGCTATTTACCAAT 180  
QY 222 CTGAAAGTATGCTGGGAAATTTCAAGAGAGAGTCAAGAGAGGCTCTGAACAGCAATT 281  
Db 181 CTGAAAGTATGCTGGGAAATTTCAAGAGAGAGTCAAGAGAGGCTCTGAACAGCAATT 240  
QY 282 AACCCATTCAGATCTGCGCTACTGAACTACATAGCCGCCACAGATGACCTC 341  
Db 241 AACCCATTCAGATCTGCGCTACTGAACTACATAGCCGCCACAGATGACCTC 300  
QY 342 ATGCTATCAAGCTGGCTTAAGCTGCAATGCTCAATCCAAAGTCCAGCCCTTCCCTC 401  
Db 301 ATGCTATCAAGCTGGCTTAAGCTGCAATGCTCAATCCAAAGTCCAGCCCTTCCCTC 360  
QY 402 GCCACGCAAGTGTACGGGCGAGGCACTGTCTGTCTCTCAGGCTTGAAGTGAAGCCAA 461  
Db 361 GCCACGCAAGTGTACGGGCGAGGCACTGTCTGTCTCTCAGGCTTGAAGTGAAGCCAA 420  
QY 462 GAAACAGTGGCCGACACCTGACTTGGCGGCAAACTGGAAGGCCCGGTGATGCTGAT 521  
Db 421 GAAACAGTGGCCGACACCTGACTTGGCGGCAAACTGGAAGGCCCGGTGATGCTGAT 480  
QY 522 CGAAGATGCCAAAAACAGAACAGGAAAGCCACAGAAATCTTATGTGTGAATTT 581  
Db 481 CGAAGATGCCAAAAACAGAACAGGAAAGCCACAGAAATCTTATGTGTGAATTT 540  
QY 582 GTGAAGTATTCAGCGGAATTTTGGGAGAGTGGCGGCTGTCTACATCTGCAAGAAC 641  
Db 541 GTGAAGTATTCAGCGGAATTTTGGGAGAGTGGCGGCTGTCTACATCTGCAAGAAC 600  
QY 642 AAGCTCCAGGAATCGAGTGGGCACTTCAATGAGGAGGAGCTCGGCATCTTACACCAAT 701  
Db 601 AAGCTCCAGGAATCGAGTGGGCACTTCAATGAGGAGGAGCTCGGCATCTTACACCAAT 660  
QY 702 GTTTACAAATATGATTCCTGATTTGAGAACTGCTTAAGCAAGTGA 749  
Db 661 GTTTACAAATATGATTCCTGATTTGAGAACTGCTTAAGCAAGTGA 708

RESULT 9  
AX455326 708 bp DNA linear PAT 06-JUL-2002  
LOCUS Sequence 3 from Patent WO0206331.  
DEFINITION AX455326  
ACCESSION AX455326  
VERSION AX455326.1 GI:21714509  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukayoka; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 Meyers,R.  
AUTHORS 14087; a novel serine protease molecule and uses therefor  
TITLE Patent: WO 0206331-A 3 24-JAN-2002;  
JOURNAL Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..708  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 186 a 187 c 168 g 167 t  
ORIGIN  
Query Match 81.6%; Score 706.4; DB 6; Length 708;  
Best Local Similarity 99.9%; Pred. No. 3.9e-188;  
Matches 707; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 42 AAGAAATATGCTCTTATTTGGGTGCTCTGCGTGGACATTTTCTTGTGACTCATCT 101  
Db 1 AAGAAATATGCTCTTATTTGGGTGCTCTGCGTGGACATTTTCTTGTGACTCATCT 60

QY 42 ATGAATATGCTTTCTATTGGGTCCTCGCTGGAGACATTTTCTTCTGCTCATCT 101  
 DB 1 ATGAATATGCTTTCTATTGGGTCCTCGCTGGAGACATTTTCTTCTGCTCATCT 60  
 QY 102 GTTCAAGAAAGAACCTCTGCTCTATTTTGGTGTACTCAAGTCTCATCTCAACCCCTGT 161  
 DB 61 GTTCAAGAAAGAACCTCTGCTCTATTTTGGTGTACTCAAGTCTCATCTCAACCCCTGT 120  
 QY 169 GTGGGGCTCTCATCAACCCAGCTGGGTGCTGGCCCGCCTCACTGCTATTTTACCAAT 221  
 DB 121 GTGGGGCTCTCATCAACCCAGCTGGGTGCTGGCCCGCCTCACTGCTATTTTACCAAT 180  
 QY 222 CTGAAGTGTATCTGGGAATTTTCAAGACAGAGTCAAGACGGTACTGAACAGCAAT 281  
 DB 181 CTGAAGTGTATCTGGGAATTTTCAAGACAGAGTCAAGACGGTACTGAACAGCAAT 240  
 QY 282 AACCCCATTCAGATCGTCCGCTACTGGAACCTACAGTATAGCGCCACAGAGTACCTC 341  
 DB 241 AACCCCATTCAGATCGTCCGCTACTGGAACCTACAGTATAGCGCCACAGAGTACCTC 300  
 QY 342 ATGCTCATCAAGCTGGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
 DB 301 ATGCTCATCAAGCTGGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTC 360  
 QY 402 GCCACCAACCAATGTCAAGCCAGCACTGTCTGCTACTCTCAAGTGTGAGTGAAGCCAA 461  
 DB 361 GCCACCAACCAATGTCAAGCCAGCACTGTCTGCTACTCTCAAGTGTGAGTGAAGCCAA 420  
 QY 462 GAAAAAGTGGCGGACACCTCTGACTTGGCGGAGAACCTGAGAGCCCGTGTATGTGAT 521  
 DB 421 GAAAAAGTGGCGGACACCTCTGACTTGGCGGAGAACCTGAGAGCCCGTGTATGTGAT 480  
 QY 522 CGAGAAATGCGCAAAAAACAAGAAAAAGCAAGAAATTTCTTAATGTGAAATTT 581  
 DB 481 CGAGAAATGCGCAAAAAACAAGAAAAAGCAAGAAATTTCTTAATGTGAAATTT 540  
 QY 582 GTGAAGATTTTCAAGCGGAATTTTGGGAGGTGGCCGTTGCTACTGCTCAATCTGCAAGAC 641  
 DB 541 GTGAAGATTTTCAAGCGGAATTTTGGGAGGTGGCCGTTGCTACTGCTCAATCTGCAAGAC 600  
 QY 642 AAGCTTCAGAGGAATCGAGGTGGGCACTTCATGAGAGGGAAGTGGCATCTCAACCAAT 701  
 DB 601 AAGCTTCAGAGGAATCGAGGTGGGCACTTCATGAGAGGGAAGTGGCATCTCAACCAAT 660  
 QY 702 GTTACAAATATGTATCTGGAATGGAACACTGCTAAGACAGATGA 749  
 DB 661 GTTACAAATATGTATCTGGAATGGAACACTGCTAAGACAGATGA 708  
 RESULT 10  
 AX350481 721 bp DNA linear PART 06-FEB-2002  
 LOCUS Sequence 106 from Patent WO0181578.  
 DEFINITION AX350481  
 ACCESSION AX350481 GI:18616087  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 Vernet,C.A., Fernandes,E.R., Gerlach,V., Shinkets,R.A.,  
 Malyanfar,U.M., Boldog,F.L., Zernhusen,B.D., Spytek,K.A.,  
 Majumder,K., Tchertnev,V.T., Padigaru,M., Patuzaj,M.,  
 Burgess,C.E., Gangoli,B.A., Smithson,G., Rastelli,L.,  
 Macdonald,J.R., Taupier,R.J., Grose,W.M. and Alsobrook,J.P.  
 Novel protease and nucleic acids encoding same  
 Patent: WO 0181578-A 106 01-NOV-2001;  
 TITLE Curagen Corporation (US)  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source 1..721  
 /organism="Homo sapiens"

BASE COUNT 193 a 191 c 169 g 168 t  
 ORIGIN  
 Query Match 81.3%; Score 703.8; DB 6; Length 721;  
 Best Local Similarity 99.6%; Pred. No. 2,1e-187;  
 Matches 716; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
 QY 39 ACCATGAATATGCTTTCTATTGGGTCCTCGCTGGAGACATTTTCTTCTGCTCATCT 98  
 DB 2 ACCATGAATATGCTTTCTATTGGGTCCTCGCTGGAGACATTTTCTTCTGCTCATCT 61  
 QY 99 TCTGTTCAAGAAAGAACCTCTGCTCTATTTTGGTGTACTCAAGTCTCATCTCAACCC 158  
 DB 62 TCTGTTCAAGAAAGAACCTCTGCTCTATTTTGGTGTACTCAAGTCTCATCTCAACCC 121  
 QY 159 TGTGTGGGGCTCTCATCAACCCAGCTGGGTGCTGGCCCGCCTCACTGCTATTTTACCA 218  
 DB 122 TGTGTGGGGCTCTCATCAACCCAGCTGGGTGCTGGCCCGCCTCACTGCTATTTTACCA 181  
 QY 219 AATCTGAAGTGTATCTGGGAATTTTCAAGACAGAGTCAAGACGGTACTGAACAGACA 278  
 DB 182 AATCTGAAGTGTATCTGGGAATTTTCAAGACAGAGTCAAGACGGTACTGAACAGACA 241  
 QY 279 ATTAACCCCATTCAGATCGTCCGCTACTGGAACCTACAGTATAGCGCCACAGAGTAC 338  
 DB 242 ATTAACCCCATTCAGATCGTCCGCTACTGGAACCTACAGTATAGCGCCACAGAGTAC 301  
 QY 339 CTCAATGCTCATCAAGCTGGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCC 398  
 DB 302 CTCAATGCTCATCAAGCTGGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTACC 361  
 QY 399 CTGCGCACCAACCAATGTCAAGCCAGCACTGTCTGCTACTCTCAAGTGTGAGTGAAG 458  
 DB 362 CTGCGCACCAACCAATGTCAAGCCAGCACTGTCTGCTACTCTCAAGTGTGAGTGAAG 421  
 QY 459 CAAGAAACAGTGGCGGACACCTGACTTGGCGGAGAACCTGAGAGCCCGTGTATGTCT 518  
 DB 422 CAAGAAACAGTGGCGGACACCTGACTTGGCGGAGAACCTGAGAGCCCGTGTATGTCT 481  
 QY 519 GATCAGAAATGCC-AAAAACAGAAACAAGAAAAAGCCACAGAAATTTCTTAATGTGGA 577  
 DB 482 GATCAGAAATGCCCAAAAAACAAGAAAAAGCCACAGAAATTTCTTAATGTGGA 541  
 QY 578 ATTTGTGAAGATTTTCAAGCGGAATTTTGGGAGGTGGCCGTTGCTACTGCTCATCTGCAA 637  
 DB 542 ATTTGTGAAGATTTTCAAGCGGAATTTTGGGAGGTGGCCGTTGCTACTGCTCATCTGCAA 601  
 QY 638 AGACAAGCTCCAGGGAATCGAGGTGGGCACTTCATGAGAGGGAAGTGGCATCTACAC 697  
 DB 602 AGACAAGCTCCAGGGAATCGAGGTGGGCACTTCATGAGAGGGAAGTGGCATCTACAC 661  
 QY 698 CAATGTTTACAAATATGTATCTGGAATGGAACACTGCTAAGACAGATGAAGACCCCTA 756  
 DB 662 CAATGTTTACAAATATGTATCTGGAATGGAACACTGCTAAGACAGATGAAGACCCCTA 720  
 RESULT 11  
 AX360087 798 bp DNA linear PART 13-FEB-2002  
 LOCUS Sequence 43 from Patent WO0200860.  
 DEFINITION AX360087  
 ACCESSION AX360087  
 VERSION AX360087.1 GI:18675713  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 Plozman,G., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and  
 Charydzak,G.  
 TITLE Novel proteases

JOURNAL Patent: WO 0200860-A 43 03-JAN-2002;  
Sugen, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..798  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 212 a 212 c 191 g 183 t  
ORIGIN

Query Match 70.0%; Score 606.4; DB 6; Length 798;  
Best Local Similarity 88.4%; Pred. No. 6.9e-160;  
Matches 707; Conservative 0; Mismatches 1; Indels 90; Gaps 1;

QY 42 ATGAATATGCTCTTCTATTTGGGTGTCCTCGCTGGACATTTTCTTCTGCTGACTATCT 101  
DB 1 ATGAATATGCTCTTCTATTTGGGTGTCCTCGCTGGACATTTTCTTCTGCTGACTATCT 60  
QY 102 GTTCAGAAAGAACCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT 161  
DB 61 GTTCAGAAAGAACCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT 120  
QY 162 GTGGGGCTCTCTATCAAAACCCAGTGGTGTCTGGCCCCACTGCTCTATTTACCAAT 221  
DB 121 GTGGGGCTCTCTATCAAAACCCAGTGGTGTCTGGCCCCACTGCTCTATTTACCAAT 180  
QY 222 CTGAAGATGATGCTGGAAATTTCAAGACAGAGTCAAGACGCTACTGAACAGCAAT 281  
DB 181 CTGAAGATGATGCTGGAAATTTCAAGACAGAGTCAAGACGCTACTGAACAGCAAT 240  
QY 282 AACCCCATTCAGATCGTCCGCTACTGGAATACAGTACGCGCCCAAGATGACTC 341  
DB 241 AACCCCATTCAGATCGTCCGCTACTGGAATACAGTACGCGCCCAAGATGACTC 300  
QY 342 ATGCTATCAAGTGGTGTAGCCCTGCACTGCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
DB 301 ATGCTATCAAGTGGTGTAGCCCTGCACTGCTCAATCCCAAGTCCAGCCCTTCCCTC 360  
QY 402 GCCACCAACCAATGTCAGGACGAGCACTGTCTGTCTACTCTCAGGTTTGAAGTGAAGCA 461  
DB 361 GCCACCAACCAATGTCAGGACGAGCACTGTCTGTCTACTCTCAGGTTTGAAGTGAAGCA 420  
QY 462 GAAACAGT----- 470  
DB 421 GAAACAGTGGGCTTTGGCAGCTGAGCAACAGCCATCTGACAGAGGCCA 480  
QY 471 -----GCCGACACCTGACTTGGCG 491  
DB 481 GCCATTCCTGATTGGCAGAGACCAATTCACATGAACAGGCCGACACCTGACTTGGCG 540  
QY 492 CAGAACCTGAGAGCCCCCGTATGTCTGATCGAGATGCGCAAAAACAGAAACAGAGAAA 551  
DB 541 CAGAACCTGAGAGCCCCCGTATGTCTGATCGAGATGCGCAAAAACAGAAACAGAGAAA 600  
QY 552 AGCCACAGAAATCTTATGTGTGAATTTGTGAAGTATTCAGCCGAATTTTGGGAG 611  
DB 601 AGCCACAGAAATCTTATGTGTGAATTTGTGAAGTATTCAGCCGAATTTTGGGAG 660  
QY 612 GTGGCCGTCTACTGTCTATCTGCAAAAGACAGTCTCAGGGAATCGAGGTGGGCACTTC 671  
DB 661 GTGGCCGTCTACTGTCTATCTGCAAAAGACAGTCTCAGGGAATCGAGGTGGGCACTTC 720  
QY 672 ATGGAGAGGGAGCGTGGCATCTACCAAGTGTTAACAATATGATCTGGAATTGGAAC 721  
DB 721 ATGGAGAGGGAGCGTGGCATCTACCAAGTGTTAACAATATGATCTGGAATTGGAAC 780  
QY 732 ACTGCTAAGGACAAAGTGA 749  
DB 781 ACTGCTAAGGACAAAGTGA 798

RESULT 12  
AK097648

LOCUS AK097648 1855 bp mRNA linear PRI 15-JUL-2002  
DEFINITION Homo sapiens cDNA FLJ40329 fis, clone TEST12031418, weakly similar  
to TRYPsin I-P1 PRECURSOR (EC 3.4.21.4).  
ACCESSION AK097648  
VERSION AK097648.1 GI:21757486  
KEYWORDS oligo capping, fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS 1 Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S., Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hozuta, T., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuna, M., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nishihara, K., Maehno, Y., Nagai, K. and Isogai, T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1855)  
AUTHORS Isogai, T. and Yamamoto, J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, Flj Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
COMMENT (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
source 1..1855  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TEST12031418"  
/isue\_type="testis"  
/clone\_lib="TEST12"  
/note="Cloning vector: pME18SFL3"  
BASE COUNT 564 a 438 c 408 g 445 t  
ORIGIN

Query Match 60.9%; Score 527; DB 9; Length 1855;  
Best Local Similarity 87.4%; Pred. No. 2e-137;  
Matches 627; Conservative 0; Mismatches 0; Indels 90; Gaps 1;

QY 218 AATCTGAAGTATGCTGGGAAATTTCAAGACAGAGTCAGACGCTGTAACAGAC 277  
DB 1123 AATCTGAAGTATGCTGGGAAATTTCAAGACAGAGTCAGACGCTGTAACAGAC 1182  
QY 278 AATTAACCCATTCAGATCGTCCGCTACTGGAATCAAGTACAGGCCCCCAGAGATGA 317  
DB 1183 AATTAACCCATTCAGATCGTCCGCTACTGGAATCAAGTACAGGCCCCCAGAGATGA 1242  
QY 338 CCTCATGCTCATCAAGCTGGCTAAGCTCGCATGCTCAATCCCAAGTCCAGCCCTTCC 397  
DB 1243 CCTCATGCTCATCAAGCTGGCTAAGCTCGCATGCTCAATCCCAAGTCCAGCCCTTCC 1302  
QY 398 CCTGCGCACCAACAAATGTCAGGCAAGCACTGTCTGTCTACTCTCAGGTTTGGACTGGAG 457  
DB 1303 CCTGCGCACCAACAAATGTCAGGCAAGCACTGTCTGTCTACTCTCAGGTTTGGACTGGAG 1362  
QY 458 CCAAGAAAACAGT----- 470  
DB 1363 CCAAGAAAACAGTGGGCTTTGGAGCTGAGACCAAGGCCATCTGACTGACAGAG 1422  
QY 471 -----GCCGACACCTGACTT 487

Db 1423 CCCAGCCATTCTGATTGGAGAGACAAATTCATGAAAGGCCGACACCTGTACTT 1482  
Qy 488 GCGGCAAGAACTTGAGAGGCCCCCGTGAATGTCTGATCGAATGCGAAAGAGAGAGG 547  
Db 1483 GCGGCAAGAACTTGAGAGGCCCCCGTGAATGTCTGATCGAATGCGAAAGAGAGAGG 1542  
Qy 548 AAAAGCCAGAGAAATTCCTTATGTGTGAATTTGTGAAGATTCAGCCGAATTTTGG 607  
Db 1543 AAAAGCCAGAGAAATTCCTTATGTGTGAATTTGTGAAGATTCAGCCGAATTTTGG 1602  
Qy 608 GAGGAGGCGCGTGTCTACTGTCTGCAAGAACAGCTCCAGGGAATGAGGTGGGCA 667  
Db 1603 GAGGAGGCGCGTGTCTACTGTCTGCAAGAACAGCTCCAGGGAATGAGGTGGGCA 1662  
Qy 668 CTTCATGGAGGAGGAGCGTGGCATCTACACCAATGTTTAAATATGATCTCGAATGA 727  
Db 1663 CTTCATGGAGGAGGAGCGTGGCATCTACACCAATGTTTAAATATGATCTCGAATGA 1722  
Qy 728 GAACTGTCTAAGAGACAGTGAAGACCTTCTCTCTGCAATTCAGCTGGCTTGCCA 787  
Db 1723 GAACTGTCTAAGAGACAGTGAAGACCTTCTCTCTGCAATTCAGCTGGCTTGCCA 1782  
Qy 788 TGCACTATACAGACAGATATTTTCCCTCTATTCAAATTAATTTCTCAATGAAAA 844  
Db 1783 TGCACTATACAGACAGATATTTTCCCTCTATTCAAATTAATTTCTCAATGAAAA 1839

RESULT 13  
AR240729 AR240729 705 bp DNA 1linear PAT 20-DEC-2002  
LOCUS Sequence 3 from patent US 6468776.  
DEFINITION AR240729  
ACCESSION AR240729  
VERSION AR240729.1 GI:27285917  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 705)  
AUTHORS Konklin, D.C.  
TITLE Human serine protease  
JOURNAL Patent: US 6468776-A 3 22-OCT-2002;  
FEATURES Location/Qualifiers  
source 1..705  
BASE COUNT 133 a 68 c 116 g 102 t 286 others  
ORIGIN

Query Match 59.0%; Score 511.2; DB 6; Length 705;  
Best Local Similarity 59.3%; Pred. No. 5,4e-133;  
Matches 418; Conservative 163; Mismatches 124; Indels 0; Gaps 0;

Qy 42 ATGAATATATGCTCTTATTTGGGTGCTCGCTGGAGCAATTTTCTTGTGACTATCT 101  
Db 1 ATGAATATATGCTCTTATTTGGGTGCTCGCTGGAGCAATTTTCTTGTGACTATCT 60  
Qy 102 GTTCAGAAAGAGACCTGCTCCCTATTTGGTGTACTCAAGCTCACTTCAACCCCTGT 161  
Db 61 GTTCAGAAAGAGACCTGCTCCCTATTTGGTGTACTCAAGCTCACTTCAACCCCTGT 120  
Qy 162 GTGGGGGCTCCTCAACCAAGCTGGGTGCTGGCCCCAGCTCACTGCTATTTACCAAT 221  
Db 121 GTGGGGGCTCCTCAACCAAGCTGGGTGCTGGCCCCAGCTCACTGCTATTTACCAAT 180  
Qy 222 CTGAAGATGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 281  
Db 181 CTGAAGATGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 240  
Qy 282 AATCCATTTGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 341  
Db 241 AATCCATTTGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 300  
Qy 342 ATGCTCATCAGGCTGCTAAGCTGCTCAATCCCAAGTCCAGGCCCTTCCCTC 401  
Db 401 ATGCTCATCAGGCTGCTAAGCTGCTCAATCCCAAGTCCAGGCCCTTCCCTC 461

Db 301 ATGTTNATHAARYTNGCNAARCCNGCATNTATTAAYCCNABGTNCARCCNTNACNTN 360  
Qy 402 GCACACCAATATGTCAGGCGAGGCACTGTCTGATCTACGATTTGGACTGAGCCAA 461  
Db 361 GCACACCAATATGTCAGGCGAGGCACTGTCTGATCTACGATTTGGACTGAGCCAA 420  
Qy 462 GAAACAGTGGCGGACACCTGATCTGCGGAGAACCTGAGAGGCCCGTGTATGTGAT 521  
Db 421 GAAACAGTGGCGGACACCTGATCTGCGGAGAACCTGAGAGGCCCGTGTATGTGAT 480  
Qy 522 CGAAGATGCCAAAAAGAGACAGGAAAGCCAGCAATTTCTTATGTGAAATTT 581  
Db 481 CGAAGATGCCAAAAAGAGACAGGAAAGCCAGCAATTTCTTATGTGAAATTT 540  
Qy 582 GTGAAGATTTTCAAGCGGAAATTTTGGGAGGTGGCGGCTGTACTCTGATCGCAAGAC 641  
Db 541 GTGAAGATTTTCAAGCGGAAATTTTGGGAGGTGGCGGCTGTACTCTGATCGCAAGAC 600  
Qy 642 AAGCTTCAGGAAATCAGGTGGGCGGCACTTCAATGAGAGGAGCGTGGCATCTACCAAT 701  
Db 601 AAGCTTCAGGAAATCAGGTGGGCGGCACTTCAATGAGAGGAGCGTGGCATCTACCAAT 660  
Qy 702 GTTACAAATATGATCTCTGATTTGAGAACTGCTTAAGAGCAAG 746  
Db 661 GTTACAAATATGATCTCTGATTTGAGAACTGCTTAAGAGCAAG 705

RESULT 14  
AX318281 AX318281 705 bp DNA 1linear PAT 14-DEC-2001  
LOCUS Sequence 3 from Patent WO0166771.  
DEFINITION AX318281  
ACCESSION AX318281  
VERSION AX318281.1 GI:17900931  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Konklin, D.C.  
TITLE Human serine protease  
JOURNAL Patent: WO 0166771-A 3 13-SEP-2001;  
FEATURES Location/Qualifiers  
source 1..705  
BASE COUNT 133 a 68 c 116 g 102 t 286 others  
ORIGIN

Query Match 59.0%; Score 511.2; DB 6; Length 705;  
Best Local Similarity 59.3%; Pred. No. 5,4e-133;  
Matches 418; Conservative 163; Mismatches 124; Indels 0; Gaps 0;

Qy 42 ATGAATATATGCTCTTATTTGGGTGCTCGCTGGAGCAATTTTCTTGTGACTATCT 101  
Db 1 ATGAATATATGCTCTTATTTGGGTGCTCGCTGGAGCAATTTTCTTGTGACTATCT 60  
Qy 102 GTTCAGAAAGAGACCTGCTCCCTATTTGGTGTACTCAAGCTCACTTCAACCCCTGT 161  
Db 61 GTTCAGAAAGAGACCTGCTCCCTATTTGGTGTACTCAAGCTCACTTCAACCCCTGT 120  
Qy 162 GTGGGGGCTCCTCAACCAAGCTGGGTGCTGGCCCCAGCTCACTGCTATTTACCAAT 221  
Db 121 GTGGGGGCTCCTCAACCAAGCTGGGTGCTGGCCCCAGCTCACTGCTATTTACCAAT 180  
Qy 222 CTGAAGATGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 281  
Db 181 CTGAAGATGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 240  
Qy 282 AATCCATTTGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 341  
Db 341 AATCCATTTGATGCTGGGAAATTTCAAGAGAGTCAAGCGTACTGAACAGCAAT 300









PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR P-PSDB; AAM39438.  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 XX such as central nervous system injuries -  
 XX  
 PS Claim 1; SEQ ID NO 797; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA15798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 866 BP; 233 A; 227 C; 197 G; 209 T; 0 other;

Query Match 100.0%; Score 866; DB 22; Length 866;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-243;  
 Matches 866; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACTTACTCTCCCTGAGCTAAAGGGGGAAGAGCTGATCCACCAATATATGCTCTTATTT 60  
 DB 1 GGCACTTACTCTCCCTGAGCTAAAGGGGGAAGAGCTGATCCACCAATATATGCTCTTATTT 60  
 QY 61 TGGGTGCTCGCGGGGCAATTTTCTTGCTGATCATCTGTTTCAAGAAAGAACCCCTG 120  
 DB 61 TGGGTGCTCGCGGGGCAATTTTCTTGCTGATCATCTGTTTCAAGAAAGAACCCCTG 120  
 QY 121 CTCCTATTTGTTGTTACTCAAGTCTCACTTCAACCCCTGTGTGGCGTCTCATCAAC 180  
 DB 121 CTCCTATTTGTTGTTACTCAAGTCTCACTTCAACCCCTGTGTGGCGTCTCATCAAC 180  
 QY 181 CGAGCTGGGTGCTGGGCCCAAGCTCACTGATTTTACCAAAATCTGAAGATGCTGGAA 240  
 DB 181 CGAGCTGGGTGCTGGGCCCAAGCTCACTGATTTTACCAAAATCTGAAGATGCTGGAA 240  
 QY 241 ATTTCAGAGCAGAGTGAAGAGGCTAATAAGCAAGAAATTAACCCCATTCAGATCTCTC 300  
 DB 241 ATTTCAGAGCAGAGTGAAGAGGCTAATAAGCAAGAAATTAACCCCATTCAGATCTCTC 300  
 QY 301 GCTACTGGAAGTCAAGTCAATGAGCCCAAGATGAGCTCATGCTCATCAAGCTGGCTA 360  
 DB 301 GCTACTGGAAGTCAAGTCAATGAGCCCAAGATGAGCTCATGCTCATCAAGCTGGCTA 360  
 QY 361 AGCCTGCGATGCTCAATCCAAAGTCAGCCCTTCCCTGCGCAACCAATGTCAGGC 420  
 DB 361 AGCCTGCGATGCTCAATCCAAAGTCAGCCCTTCCCTGCGCAACCAATGTCAGGC 420  
 QY 421 CAGGCACTGCTGCTCTCTCTCAGGTTTGAATGAGCCCAAGAAAGAGTGGCGGACACC 480  
 DB 421 CAGGCACTGCTGCTCTCTCTCAGGTTTGAATGAGCCCAAGAAAGAGTGGCGGACACC 480  
 QY 481 CTGACTTGGCGGAGAACCTGAGAGGCCCTCGATGATGATCGAATGCCAAAAACAG 540

DB 481 CTGACTTGGCGGAGAACCTGAGAGGCCCTCGATGATGATCGAATGCCAAAAACAG 540  
 QY 541 AACAAAGAAAAGCCCAAGAAATTCCTTATGTTGAAATTTGGAAAGTATTCAGCCGA 600  
 DB 541 AACAAAGAAAAGCCCAAGAAATTCCTTATGTTGAAATTTGGAAAGTATTCAGCCGA 600  
 QY 601 TTTTGGGAGGAGTGGCCGTTGCTACTGTCATCTGCAAGACAGCTCCAGGAAATCGAG 660  
 DB 601 TTTTGGGAGGAGTGGCCGTTGCTACTGTCATCTGCAAGACAGCTCCAGGAAATCGAG 660  
 QY 661 TGGGAGCACTTCATGAGAGGGGAGCTGCGCATCTACCAATGTTTCAAAATATGATCTCT 720  
 DB 661 TGGGAGCACTTCATGAGAGGGGAGCTGCGCATCTACCAATGTTTCAAAATATGATCTCT 720  
 QY 721 GGATTGGAACACTGTAAGACAAAGTGAACCCCTATCTCCCTGTGATTCACACTGGC 780  
 DB 721 GGATTGGAACACTGTAAGACAAAGTGAACCCCTATCTCCCTGTGATTCACACTGGC 780  
 QY 781 TTGCGCATGACTATACAGACAGATATTTTCCCTATTTCAAAATTAATCTCAATG 840  
 DB 781 TTGCGCATGACTATACAGACAGATATTTTCCCTATTTCAAAATTAATCTCAATG 840  
 QY 841 AAAATTTGGGAATGTAGCATCTACTAGT 866  
 DB 841 AAAATTTGGGAATGTAGCATCTACTAGT 866

RESULT 2  
 ABK15700  
 ID ABK15700 standard; cDNA; 1060 BP.  
 XX  
 AC ABK15700;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX  
 DE Human cDNA encoding novel serine protease 14087.  
 XX  
 KW Human; ss; gene; serine protease; 14087; cellular proliferative disorder;  
 KW cancer; carcinoma; leukaemia; brain disorder; meningitis; AIDS;  
 KW acquired immunodeficiency syndrome-associated myopathy; heart disorder;  
 KW multiple sclerosis; Alzheimer's disease; Parkinson's disease;  
 KW ischemic heart disease; cardiac hypertrophy; myocardial infarction;  
 KW atherosclerosis; hypertension; platelet disorder; osteoporosis;  
 KW osteomalacia; osteopenia; haematopoietic disorder; diabetes mellitus;  
 KW rheumatoid arthritis; liver disorder; viral disease; metabolic disorder;  
 KW obesity; anorexia; cachexia.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 164..871  
 FT /\*tag= a  
 FT /product= "Serine protease 14087"  
 FT sig\_peptide 164..223  
 FT /\*tag= b  
 FT mat\_peptide 224..868  
 FT /\*tag= c  
 XX  
 PN W0200206331-A2.  
 PD 24-JAN-2002.  
 XX  
 PF 18-JUL-2001; 2001WO-US23184.  
 XX  
 PR 18-JUL-2000; 2000US-219022P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Meyers R;  
 XX  
 DR WPI: 2002-188537/24.  
 DR P-PSDB; AAV76316.

XX New serine protease polypeptide for diagnosing and treating cellular  
 PT proliferation, brain, heart, bone, liver or metabolic disorders, viral  
 PT diseases and for identifying modulators

PS Claim 1; Fig 1; 105pp; English.

XX The invention relates to an isolated serine protease polypeptide, 14087,  
 CC or its naturally occurring allelic variant or fragment (and their  
 CC encoding polynucleotides). Also included are a host cell containing the  
 CC polynucleotide, an anti-14087 antibody, and a modulator of the activity  
 CC of 14087. 14087 is useful for identifying a compound which binds to or  
 CC modulates the activity of 14087. Fragments comprising at least 25  
 CC contiguous nucleotides of a nucleic acid encoding 14087 are useful  
 CC as hybridization probes or primers for identifying a nucleic acid  
 CC molecule associated with a disorder, and a subject having disorder or  
 CC at risk for developing a disorder. Assaying the ability of the compound  
 CC to modulate 14087 nucleic acid expression or 14087 polypeptide activity  
 CC is useful for identifying a compound capable of treating disorder  
 CC characterized by aberrant 14087 nucleic acid expression or polypeptide  
 CC activity. 14087 proteins are useful for modulating cellular  
 CC proliferation, differentiation, tumorigenesis, modulating an immune  
 CC response, catalyzing muscle-related reactions and modulating  
 CC proteolysis of protein substrates. 14087 protein is useful in treating  
 CC various disorders including cellular proliferative and/or  
 CC differentiative disorders (e.g. cancer, carcinoma, leukemia or  
 CC haematopoietic neoplastic disorder), brain (e.g. meningitis, acquired  
 CC immunodeficiency syndrome (AIDS)-associated myopathy, multiple  
 CC sclerosis, Alzheimer's disease, Parkinson's disease), heart (e.g.  
 CC ischaemic heart disease, cardiac hypertrophy, myocardial  
 CC infarction), blood vessel (e.g. atherosclerosis, hypertension, Kawasaki  
 CC syndrome), and platelet disorders, as well as disorders associated with  
 CC bone metabolism (e.g. osteoporosis, osteomalacia, osteopenia, tropical  
 CC sprue, rickets), haematopoietic (e.g. diabetes mellitus, rheumatoid  
 CC arthritis, autoimmune thyroiditis), liver disorders, viral diseases and  
 CC pain or metabolic disorders (e.g. obesity, anorexia, cachexia) and  
 CC many other diseases and disorders given in the specification. The  
 CC present sequence is the cDNA encoding serine protease 14087.

XX Sequence 1060 BP, 309 A, 259 C, 237 G, 251 T, 4 other;

XX Query Match 99.6%; Score 862.8; DB 24; Length 1060;  
 XX Best Local Similarity 99.8%; Pred. No. 6,9e-242;  
 XX Matches 864; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGCACCTTACTCCCTGAGCTAAGGGGGAAGAGCTGATCACCATGAATAATGCTTCTATT 60  
 DB 123 GACACTTAATCTCCCTGAGCTAAGGGGGAAGAGCTGATCACCATGAATAATGCTTCTATT 182  
 QY 61 TGGGTGCTCTCGCTGGAGCACTTTTCTTCTGCTACATCTGTTCAGAAAGAACCTTG 120  
 DB 183 TGGGTGCTCTCGCTGGAGCACTTTTCTTCTGCTACATCTGTTCAGAAAGAACCTTG 242  
 QY 121 CTCCTAATTGAGTGTACCTCAAGTCACTTAACCCCTGTGAGGCGTCTCATCAAC 180  
 DB 243 CTCCTAATTGAGTGTACCTCAAGTCACTTAACCCCTGTGAGGCGTCTCATCAAC 302  
 QY 181 CCAGCTGGGTGCTGGCCCGCCAGCTCACTGCTATTTCACAAATCTGAAGTATGCTGGAA 240  
 DB 303 CCAGCTGGGTGCTGGCCCGCCAGCTCACTGCTATTTCACAAATCTGAAGTATGCTGGAA 362  
 QY 241 ATTTCAAGAGCAGATCAAGAGCGGTACTGAAACAAATTAACCCCATTCAGATGCTCC 300  
 DB 363 ATTTCAAGAGCAGATCAAGAGCGGTACTGAAACAAATTAACCCCATTCAGATGCTCC 422  
 QY 301 GCTACTGGAATCACTACATGATAGGCGCCACAGAGATGACCTCATGCTCATCAAGCTGCTA 360  
 DB 423 GCTACTGGAATCACTACATGATAGGCGCCACAGAGATGACCTCATGCTCATCAAGCTGCTA 482  
 QY 361 AGCTTGCATGCTCAATCCCAAGATCCAGCCCTTCCCTCGGCACACCAATATGTCAGGC 420  
 DB 483 AGCTTGCATGCTCAATCCCAAGATCCAGCCCTTCCCTCGGCACACCAATATGTCAGGC 542

QY 421 CAGGACCTGTCTGTACTCTCAGGTTTGAGCTGAGCCAGAAAAACAGTGGCCGACACC 480  
 DB 543 CAGGACCTGTCTGTACTCTCAGGTTTGAGCTGAGCCAGAAAAACAGTGGCCGACACC 602  
 QY 481 CTGACTTGGCGAGAACTGTGAGGCCCCCGGTATGTCTCATGTGAAGATCCAAAAACAG 540  
 DB 603 CTGACTTGGCGAGAACTGTGAGGCCCCCGGTATGTCTCATGTGAAGATCCAAAAACAG 662  
 QY 541 AACAGAGAAAAAGCCACAGAAATTCCTATGTGTGAAATTTGTGAAGATTCAGCCGAA 600  
 DB 663 AACAGAGAAAAAGCCACAGAAATTCCTATGTGTGAAATTTGTGAAGATTCAGCCGAA 722  
 QY 601 TTTTGGGAGGTGGCCGCTTGTCTATCTCATCTGCAAGACAGCTCCAGGAAATGAGG 660  
 DB 723 TTTTGGGAGGTGGCCGCTTGTCTATCTCATCTGCAAGACAGCTCCAGGAAATGAGG 782  
 QY 661 TGGGACATTCATGTGAGGAGGAGCGTGGCATCTACACCAATGTTTCAAAATATGTTCT 720  
 DB 783 TGGGACATTCATGTGAGGAGGAGCGTGGCATCTACACCAATGTTTCAAAATATGTTCT 842  
 QY 721 GGATTGAGAGACGTGCTAAGCAAGTGAAGACCTTCTCCCTGCAATTCACCTGGC 780  
 DB 843 GGATTGAGAGACGTGCTAAGCAAGTGAAGACCTTCTCCCTGCAATTCACCTGGC 902  
 QY 781 TCTGCATGACTATACAGACAGATATTTTCTCTATTCACAAATTAATTCACCAATG 840  
 DB 903 TCTGCATGACTATACAGACAGATATTTTCTCTATTCACAAATTAATTCACCAATG 962  
 QY 841 AAAATTTGGGAATGTAGCATCTACTAGT 866  
 DB 963 AAAATTTGGGAATGTAGCATCTACTAGT 988

RESULT 3  
 ABK12889  
 ID ABK12889 standard; cDNA; 890 BP.

XX ABK12889;

XX 09-APR-2002 (first entry)

XX Human protease PRS-6 cDNA sequence.

XX Human; protease; PRS; gastrointestinal; Crohn's disease; cancer;  
 KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;  
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;  
 KW cell proliferative disorder; developmental disorder; epilepsy;  
 KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;  
 KW reproductive disorder; endometriosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 63..770  
 FT /\*tag= a  
 FT /product= "Human protease PRS-6"

XX WO200198468-A2.

XX 27-DEC-2001.

XX 13-JUN-2001; 2001WO-US19178.

XX 16-JUN-2000; 2000US-212336P.

XX 22-JUN-2000; 2000US-213955P.

XX 29-JUN-2000; 2000US-215395P.

XX 07-JUL-2000; 2000US-216821P.

XX 14-JUL-2000; 2000US-218946P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Elliott VS, Gandhi AR, Lai P, Au-Young J, Tribouley CM;  
 PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hatfield A, Khan FA;

PI Walia NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT,  
PI Aaimai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L,  
PI Kallik DA;  
XX  
XX WPI: 2002-090437/12.  
DR P-PSDB; AAU074746.  
XX  
XX Twenty one human proteases (referred to as PRS-1 to PRS-21), useful  
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.  
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell  
PT proliferative (e.g. cancer) disorders -  
XX  
XX Claim 5; Page 164; 177pp; English.  
XX  
XX The present invention relates to twenty one human proteases,  
CC referred to as PRS-1 to PRS-21. The PRS polynucleotides and  
CC polypeptides of the invention are useful in the diagnosis, treatment and  
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and  
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and  
CC myocardial infarction, autoimmune/inflammatory e.g. acquired  
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell  
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker  
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.  
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and  
CC endometriosis disorders. Numerous other examples of each disorder are  
CC given in the specification. The present nucleic acid sequence encodes  
CC the human protease PRS-6 protein of the invention.  
XX  
XX Sequence 890 BP; 249 A; 230 C; 203 G; 208 T; 0 other;

Query Match 99.1%; Score 857.8; DB 24; Length 890;  
Best Local Similarity 99.8%; Pred. No. 1.8e-240;  
Matches 859; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGCACTTACTCTCCCTGAGCTAAGGGGGAAGAGCTGATCCAGTAAATATGCTTCTATT 60  
DB 22 GGCACTTACTCTCCCTGAGCTAAGGGGGAAGAGCTGATCCAGTAAATATGCTTCTATT 81  
QY 61 TGGGTGCTCGCTGGGGAATTTTCTTGTGCTGATCTGATCTTCAAGAAAGACCTG 120  
DB 82 TGGGTGCTCGCTGGGGAATTTTCTTGTGCTGATCTGATCTTCAAGAAAGACCTG 141  
QY 121 CTCCTATTTGTGTACTCAAGTCTCACTTCAACCCCTGTGGGCTCTCATCAAC 180  
DB 142 CTCCTATTTGTGTACTCAAGTCTCACTTCAACCCCTGTGGGCTCTCATCAAC 201  
QY 181 CGAGCTGGGTGCTGGCCCGAGCTCACTGATTTACCAATCTGAAGTATGCTGGAA 240  
DB 202 CGAGCTGGGTGCTGGCCCGAGCTCACTGATTTACCAATCTGAAGTATGCTGGAA 261  
QY 241 ATTTCAAGACAGAGTGAAGAGGTACTGAACGAATTAACCCGATTCAGATGCTGC 300  
DB 262 ATTTCAAGACAGAGTGAAGAGGTACTGAACGAATTAACCCGATTCAGATGCTGC 321  
QY 301 GCTACTGGAATCACTAGTCAATAGCCGCCACAGATGACTCATGCTCAAGCTGCTA 360  
DB 322 GCTACTGGAATCACTAGTCAATAGCCGCCACAGATGACTCATGCTCAAGCTGCTA 381  
QY 361 AGCTGCTCATGCTCATATCCCAAGTCCAGCCCTTCCCTGCGACCAACAATGTCAGGC 420  
DB 382 AGCTGCTCATGCTCATATCCCAAGTCCAGCCCTTCCCTGCGACCAACAATGTCAGGC 441  
QY 421 CAGGACCTGTCTGTCTCTCTCAGGCTTGTGACTGGAACCAAGAAAGTGGCGACACC 480  
DB 442 CAGGACCTGTCTGTCTCTCTCAGGCTTGTGACTGGAACCAAGAAAGTGGCGACACC 501  
QY 481 CTGACTTGGCGAGAACTGAGGCCCCCGTGAATGTGATTCGAGATGCAAAAAACAG 540  
DB 502 CTGACTTGGCGAGAACTGAGGCCCCCGTGAATGTGATTCGAGATGCAAAAAACAG 561  
QY 541 AACAGGAAAAAGCCACAGGAATTCCTTATGTGAAATTTGAAAGTATTCAGCCGAA 600  
DB 562 AACAGGAAAAAGCCACAGGAATTCCTTATGTGAAATTTGAAAGTATTCAGCCGAA 621

QY 601 TTTTGGGGAGGTGCGCTGTGCTACTGTCATCTGCAAGACAGCTCCAGGAATCGAG 660  
DB 622 TTTTGGGGAGGTGCGCTGTGCTACTGTCATCTGCAAGACAGCTCCAGGAATCGAG 681  
QY 661 TGGGGCACTTCATGGAAGGGGACGTGGGCACTTACCAACAATGTTTCAAAATATGATCT 720  
DB 682 TGGGGCACTTCATGGAAGGGGACGTGGGCACTTACCAACAATGTTTCAAAATATGATCT 741  
QY 721 GAAATGAAACACTGTAAGGACAAGTGAACCTTCTCCCTGCAATTCACCTGCG 780  
DB 742 GAAATGAAACACTGTAAGGACAAGTGAACCTTCTCCCTGCAATTCACCTGCG 801  
QY 781 TCGCCATGACATATCAAGCAGATATTTTCCCTATTCAAATAATCTCAATG 840  
DB 802 TCGCCATGACATATCAAGCAGATATTTTCCCTATTCAAATAATCTCAATG 861  
QY 841 AAAATTTGGGAATGTACATA 861  
DB 862 AAAATTTGGGAATGTACATA 882

RESULT 4  
ID AAS12970 standard; DNA; 865 BP.  
AC AAS12970;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX Human Ztryp3 serine protease DNA.  
DE  
XX Human: Ztryp3; serine protease; asthma; vascular function; inflammation;  
KW gene therapy; stroke; testicular function; spermatogenesis; hemostatic;  
KW mass spectrometry; circular dichroism; X-ray crystallography; ds;  
KW nuclear magnetic resonance spectroscopy; antiasthmatic; antiinflammatory;  
KW cerebroprotective.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 42..749  
FT CDS /tag= a  
FT /product= "Human Ztryp3"  
FT sig\_peptide 42..98  
FT /tag= b  
FT mat\_peptide 99..746  
FT /tag= c  
FT /product= "Mature human Ztryp3"  
XX  
XX W0200166771-A2.  
XX  
XX 13-SEP-2001.  
PD  
XX 28-FEB-2001; 2001WO-US06432.  
PF  
XX 03-MAR-2000; 2000US-0518387.  
PR  
XX (ZYMO) ZYMOGENETICS INC.  
PA Conklin DC;  
XX  
XX WPI: 2001-589946/66.  
DR P-PSDB; AAU07697.  
XX  
XX Novel Ztryp3 polypeptides and polynucleotides useful in the treatment  
PT of asthma, vascular disorders including stroke, inflammation and  
PT testicular function -  
XX  
XX Claim 4; Page 79-81; 82pp; English.  
PS  
XX The invention relates to an isolated human Ztryp3 polypeptide, a member  
CC of the serine protease family. Ztryp3 polypeptides and their associated

CC polynucleotides are useful in diagnosis, therapy and industry and are  
CC used as targets for identifying modulators, preferably inhibitors of  
CC serine protease activity. The sequences are useful in the treatment of  
CC asthma, vascular function such as stroke, inflammation and testicular  
CC function (by modulating spermatogenesis). Ztryp3 proteins can be used for  
CC identifying peptide cleavage sites and for coupling amino and carboxy  
CC terminal tags. The polypeptides are also useful to teach analytical  
CC skills such as mass spectrometry, circular dichroism, X-ray  
CC crystallography and nuclear magnetic resonance spectroscopy. This  
CC sequence represents DNA encoding human Ztryp3.

CC  
XX  
SQ Sequence 865 BP; 235 A; 226 C; 196 G; 208 T; 0 other;

Query Match 98.2%; Score 850.8; DB 22; Length 865;  
Best Local Similarity 99.7%; Pred. No. 2e-238;  
Matches 865; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GGCACTTACTCCTGAGCTAAGGGGGAAGAGCTGATCACCATGAATATGTTCTATT 60  
DB 1 GGCACTTACTCCTGAGCTAAGGGGGAAGAGCTGATCACCATGAATATGTTCTATT 60  
QY 61 TGGGTGTCCTCGGTGGGACATTTTCTTCTGCTCATCTCTGTTCAGAAAGAGCCCTG 120  
DB 61 TGGGTGTCCTCGGTGGGACATTTTCTTCTGCTCATCTCTGTTCAGAAAGAGCCCTG 120  
QY 121 CTCCTATTGTTGTTGACCTCAAGCTCACTTCAACCCCTGTGTGGGGCTCTCATCAAC 180  
DB 121 CTCCTATTGTTGTTGACCTCAAGCTCACTTCAACCCCTGTGTGGGGCTCTCATCAAC 180  
QY 181 CGAGCTGGGTGCTGGGCCCCAGCTCACTGATTATTCAGAAAGTATGATCTGGGA 240  
DB 181 CGAGCTGGGTGCTGGGCCCCAGCTCACTGATTATTCAGAAAGTATGATCTGGGA 240  
QY 241 ATTTCAAGAGCAGATGTCAGAGGCTACTGAACAGCAATTAACCCCATTCAGATGCTC 300  
DB 241 ATTTCAAGAGCAGATGTCAGAGGCTACTGAACAGCAATTAACCCCATTCAGATGCTC 300  
QY 301 GCTACTGGAACCTACAGTCAAGGCCCCCAAGAGTACCTCATGCTCAAGCTGGCTA 360  
DB 301 GCTACTGGAACCTACAGTCAAGGCCCCCAAGAGTACCTCATGCTCAAGCTGGCTA 360  
QY 361 AGCCTGCATGCTCAATCCCAAGTCCAGCCCTTACCTCGGCAACAGCAATGTCAGGC 420  
DB 361 AGCCTGCATGCTCAATCCCAAGTCCAGCCCTTACCTCGGCAACAGCAATGTCAGGC 420  
QY 421 CAGGCACTGCTCTTCTACTCTCAGGTTGAGCTGAGCCAAAGAAACAGTGGCCGACACC 480  
DB 421 CAGGCACTGCTCTTCTACTCTCAGGTTGAGCTGAGCCAAAGAAACAGTGGCCGACACC 480  
QY 481 CTGACTTGGGGGAGAACCTGAGAGCCCCCGTGAATGCTGATGAGAAATGCCAAAAACAG 540  
DB 481 CTGACTTGGGGGAGAACCTGAGAGCCCCCGTGAATGCTGATGAGAAATGCCAAAAACAG 540  
QY 541 AACAAAGAAAAAGCCACAGAAATTCCTTATGTGTGAAATTTGAAAGTATTCAGCCGAA 600  
DB 541 AACAAAGAAAAAGCCACAGAAATTCCTTATGTGTGAAATTTGAAAGTATTCAGCCGAA 600  
QY 601 TTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGCAAGACTTCAGAGGAATGAGG 660  
DB 601 TTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGCAAGACTTCAGAGGAATGAGG 660  
QY 661 TGGGGCACTTCATGAGAGGAGGAGCTGGGCTTACACCATGTTTACAAATATGATCTCT 720  
DB 661 TGGGGCACTTCATGAGAGGAGGAGCTGGGCTTACACCATGTTTACAAATATGATCTCT 720  
QY 721 GGATTGAAACACTGCTAAGAGCAAGTGAACCTTCTCTCCCTGCTGATTCAGTGGC 780  
DB 721 GGATTGAAACACTGCTAAGAGCAAGTGAACCTTCTCTCCCTGCTGATTCAGTGGC 780  
QY 781 TCTGCCATGAGCTATACAGCAAGTAATTTTCCCTCTATTCAAATTAATTCCTCAATG 840  
DB 781 TCTGCCATGAGCTATACAGCAAGTAATTTTCCCTCTATTCAAATTAATTCCTCAATG 840  
QY 840 TCTGCCATGAGCTATACAGCAAGTAATTTTCCCTCTATTCAAATTAATTCCTCAATG 899  
DB 840 TCTGCCATGAGCTATACAGCAAGTAATTTTCCCTCTATTCAAATTAATTCCTCAATG 899

QY 841 AAAATTGGGAATGTAGCATCTACTAGT 866  
DB 840 AAAATTGGGAATGTAGCAACTAGT 865

RESULT 5  
ID ABQ77082 standard; DNA; 865 BP.  
XX  
XX ABQ77082;  
AC  
XX  
XX 01-APR-2003 (first entry)  
DT  
XX  
XX Human serine protease Ztryp3 DNA.

DE Human serine protease; Ztryp3; gene; blood coagulation; fibrinolysis;  
XX  
XX Human; serine protease; Ztryp3; gene; blood coagulation; fibrinolysis;  
KW complement activation; fertilization; hormone production; gene therapy;  
XX  
XX somatic cell; ds.  
OS Homo sapiens.

XX  
XX Key Location/Qualifiers  
FH CDS 42..749  
FT /tag= a  
FT /product= "Ztryp3"  
XX

XX US6468776-B1.  
XX  
XX 22-OCT-2002.  
XX  
XX 28-FEB-2001; 2001US-0796110.  
XX  
XX 03-MAR-2000; 2000US-186623P.  
XX  
XX (ZYMO) ZYMOGENETICS INC.

XX Conklin DC;  
XX  
XX WPI; 2003-182091/18.  
XX  
XX P-PSDB; ABG73741.  
XX

XX Novel serine protease nucleic acid molecule useful for producing serine  
XX protease polypeptide, and as educational tool in laboratory practicum  
XX kits for courses related to genetics and molecular biology  
XX  
XX  
XX Claim 2; Column 55-58; 32pp; English.

XX  
XX This invention describes a novel human serine protease, Ztryp3. The  
XX serine protease family play a role in carefully controlled processes such  
XX as blood coagulation, fibrinolysis, complement activation, fertilization  
XX and hormone production. The products of the invention are useful for  
XX producing serine protease polypeptides; as probes or primers to clone 5'  
XX non-coding regions of Ztryp3 gene; as polymerase chain reaction primers  
XX and probes; as educational tool in laboratory practicum kits for courses  
XX related to genetics and molecular biology, protein chemistry, and  
XX antibody production and analysis; as standards or as unknowns for testing  
XX purposes; as an aid to teach a student how to prepare expression  
XX constructs for bacterial, viral, or mammalian expression, including  
XX fusion constructs, where Ztryp3 is the gene to be expressed; for  
XX determining the restriction endonuclease cleavage sites of the  
XX polynucleotides; for determining mRNA and DNA localization of Ztryp3  
XX polynucleotides in tissues (i.e., by northern and southern blotting as  
XX well as polymerase chain reaction); and for identifying related  
XX polynucleotides by nucleic acid hybridization. The products of the  
XX invention can also be used for determining mutations in the Ztryp3 gene,  
XX for determining Ztryp3 gene expression, for examining Ztryp3 gene  
XX structure, for detecting and localizing expression of Ztryp3 gene in a  
XX biological sample, in vivo diagnosis, in gene therapy, particularly  
XX in somatic cell gene therapy, and for producing transgenic mice. This  
XX sequence encodes the human serine protease Ztryp3 described in the method  
XX of the invention.

SQ Sequence 865 BP; 235 A; 226 C; 196 G; 208 T; 0 other;

Query Match 98.2%; Score 850.8; DB 25; Length 865;  
 Best Local Similarity 99.7%; Pred. No. 2e-238;  
 Matches 863; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

QY 1 GGCACCTACTCCCGAGCTAAGGGGGAAGAGCTGATCACCAATATATGCTTCTTAT 60
DB 1 GGCACCTACTCCCGAGCTAAGGGGGAAGAGCTGATCACCAATATATGCTTCTTAT 60
QY 61 TGGGTGCTCTGCTGGGACATTTTCTTCTGATCATCTGTTACAGAAAGAACCTTG 120
DB 61 TGGGTGCTCTGCTGGGACATTTTCTTCTGATCATCTGTTACAGAAAGAACCTTG 120
QY 121 CTCCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 180
DB 121 CTCCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 180
QY 181 CCGAGCTGGTCTGCTGGGACATTTTCTTCTGATCATCTGTTACAGAAAGAACCTTG 240
DB 181 CCGAGCTGGTCTGCTGGGACATTTTCTTCTGATCATCTGTTACAGAAAGAACCTTG 240
QY 241 ATTTCAAGAGCAGATCAGAGACGGTACTGAAAGACATTAACCCCATTCAGATGCTC 300
DB 241 ATTTCAAGAGCAGATCAGAGACGGTACTGAAAGACATTAACCCCATTCAGATGCTC 300
QY 301 GCTACTGGAACCTACAGTCTATAGCCGCCACAGATGAAGCTCATCAAGCTGCTA 360
DB 301 GCTACTGGAACCTACAGTCTATAGCCGCCACAGATGAAGCTCATCAAGCTGCTA 360
QY 361 AGCTGCTGATGCTCAATATCCCAAGTCCAGCCCTTCCCTGCGCACCAATATCAGC 420
DB 361 AGCTGCTGATGCTCAATATCCCAAGTCCAGCCCTTCCCTGCGCACCAATATCAGC 420
QY 421 CAGGCACTGCTCTGCTACTCTCAGGTTTGATGAGTGGAGCAAGAAAGTGGCCGAC 480
DB 421 CAGGCACTGCTCTGCTACTCTCAGGTTTGATGAGTGGAGCAAGAAAGTGGCCGAC 480
QY 481 CTGACTTGGCGCAGAACCTGAGGCCCCCGTATGTCGATCGAAGTGCAGAAACAG 540
DB 481 CTGACTTGGCGCAGAACCTGAGGCCCCCGTATGTCGATCGAAGTGCAGAAACAG 540
QY 541 AACAGGAAAAAGCCACAGAAATCTTATGTTGAAATTTGAAATTTGAAATTTGAA 600
DB 541 AACAGGAAAAAGCCACAGAAATCTTATGTTGAAATTTGAAATTTGAAATTTGAA 600
QY 601 TTTTGGGAGGTTGCGGCTGCTACTGATCTGCAAGCAAGCTCCAGGGAATCGAG 660
DB 601 TTTTGGGAGGTTGCGGCTGCTACTGATCTGCAAGCAAGCTCCAGGGAATCGAG 660
QY 661 TGGGGCACTTCATGGAAGGGGAGCTCGGCACTTACACCAATGTTTACAAATATG 720
DB 661 TGGGGCACTTCATGGAAGGGGAGCTCGGCACTTACACCAATGTTTACAAATATG 720
QY 721 GGATGGAACATCTGTAAGGACAGTGAACCTTACTTCCCTGCTGCAATGCGC 780
DB 721 GGATGGAACATCTGTAAGGACAGTGAACCTTACTTCCCTGCTGCAATGCGC 780
QY 781 TCTGCCATGACTATACAGAGATATTTTCCCTATTCAAAATTAATTCGAAATG 840
DB 781 TCTGCCATGACTATACAGAGATATTTTCCCTATTCAAAATTAATTCGAAATG 840
QY 841 AAAATTTGGGAATGTAGCATATAGT 866
DB 840 AAAATTTGGGAATGTAGCATATAGT 865
  
```

RESULT 6  
 ID AA160380/c  
 XX AA160380 standard; cDNA; 867 BP.  
 AC AA160380;  
 XX  
 DT 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 4369.  
 XX KW Human, neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX OS Homo sapiens.  
 XX PN MO20015312-A1.  
 XX PD 26-JUL-2001.  
 XX PE 26-DEC-2000; 2000MO-US34263.  
 XX PR 21-JAN-2000; 2000US-0488725.  
 XX PR 25-APR-2000; 2000US-0552317.  
 XX PR 09-JUL-2000; 2000US-0598042.  
 XX PR 19-JUL-2000; 2000US-0620312.  
 XX PR 03-AUG-2000; 2000US-0653450.  
 XX PR 14-SEP-2000; 2000US-0662191.  
 XX PR 19-OCT-2000; 2000US-0693036.  
 XX PR 29-NOV-2000; 2000US-0727344.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX DR WPI; 2001-442253/47.  
 XX DR P-PADB; AAM41224.  
 XX PT Novel nucleic acids and polypeptides, useful for treating disorders  
 XX PT such as central nervous system injuries -  
 XX PS Claim 1; SEQ ID NO 4369; 10078pp; English.  
 XX CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AA42213) with neurotrophic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX SQ Sequence 867 BP; 209 A; 197 C; 225 G; 236 T; 0 other;

Query Match 98.2%; Score 850.2; DB 22; Length 867;  
 Best Local Similarity 99.5%; Pred. No. 3e-238;  
 Matches 863; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

QY 1 GGCACCTACTCCCGAGCTAAGGGGGAAGAGCTGATCACCAATATATGCTTCTTAT 60
DB 867 GGCACCTACTCCCGAGCTAAGGGGGAAGAGCTGATCACCAATATATGCTTCTTAT 808
QY 61 TGGGTGCTCTGCTGGGACATTTTCTTCTGATCATCTGTTACAGAAAGAACCTTG 120
DB 807 TGGGTGCTCTGCTGGGACATTTTCTTCTGATCATCTGTTACAGAAAGAACCTTG 748
QY 121 CTCCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 180
  
```



```

Db      747 CTCCTCATTTGGTGTACTCAAGTCTCACTTCAACCCCTGTGGGGCTCTCATCAAC
Qy      181 CCAGCTGGGTGTGGCCCCAGCTCACTGTATTATTAACCAATCTGAAGTATGCTGGAA
Db      687 CCAGCTGGGTGTGGCCCCAGCTCACTGTATTATTAACCAATCTGAAGTATGCTGGAA
Qy      241 ATTTCAAGACAGAGTCAAGACGTAAGTCAAGACAAATTAACCCATTCAAGTCTCC
Db      627 ATTTCAAGACAGAGTCAAGACGTAAGTCAAGACAAATTAACCCATTCAAGTCTCC
Qy      301 GCTACTGGAATCACTCATAGAGGCCCCACAGATGACTCTATGCTCATCAAGCTGGCTA
Db      567 GCTACTGGAATCACTCATAGAGGCCCCACAGATGACTCTATGCTCATCAAGCTGGCTA
Qy      361 AGCCTGCAATGCTCAATCCCAAGTCC-AGCCCTTCCCTCCGCCCAACCAATGTCAG
Db      507 AGCCTGCAATGCTCAATCCCAAGTCCCAAGTCCCAAGTCCCAAGTCCCAAGTCC
Qy      420 CCAGGACTGTCTGTCTACTCTCAAGTTTGTGAGTGGAGCAAGAAACAGTGGCCGAC
Db      447 CCAGGACTGTCTGTCTACTCTCAAGTTTGTGAGTGGAGCAAGAAACAGTGGCCGAC
Qy      480 CTTGACTTGGGCAAGAACTGAGAGCCGCCGTGATGTTGATGAGAAATGCCAAACCA
Db      387 CTTGACTTGGGCAAGAACTGAGAGCCGCCGTGATGTTGATGAGAAATGCCAAACCA
Qy      540 GAACAGAGAAAGAAAGCAAGCAAGCAATTCCTTATGTTGAAATTTGTGAAGTATCAG
Db      327 GAACAGAGAAAGAAAGCAAGCAAGCAATTCCTTATGTTGAAATTTGTGAAGTATCAG
Qy      600 ATTTTGGGAGGTGGCCGCTTGTCTACTGTCTATGCAAGCAAGTCCAGGAAATGAG
Db      267 ATTTTGGGAGGTGGCCGCTTGTCTACTGTCTATGCAAGCAAGTCCAGGAAATGAG
Qy      660 GTGGGGCACTTCATGAGAGGGGAGCGTGGCATCTACCAATGTTTAAATATGATCC
Db      207 GTGGGGCACTTCATGAGAGGGGAGCGTGGCATCTACCAATGTTTAAATATGATCC
Qy      720 TGGATTGAGAAACCTGCTTAAGGCAAGTGAAGACCTTCTCCCTGTGATTCACCTG
Db      147 TGGATTGAGAAACCTGCTTAAGGCAAGTGAAGACCTTCTCCCTGTGATTCACCTG
Qy      780 CTTGCGCATGAGTATACAGAGATATTTCCCTCTATTCAAATTAATCTCAAAAT
Db      87 CTTGCGCATGAGTATACAGAGATATTTCCCTCTATTCAAATTAATCTCAAAAT
Qy      840 GAAATTTTGGGAAATGATAGTACTAGT 866
Db      27 GAAATTTTGGGAAATGATAGTACTAGT 1

```

RESULT 7  
ABK30293/c  
ID ABK30293 standard; cDNA; 865 BP.

XX AC ABK30293;  
XX DT 23-APR-2002 (first entry)  
XX DE Human G-protein-coupled protease #63.  
XX KM Human; ss; gene; G-protein-coupled protease; gene therapy;  
XX KM transgenic; protease mediated disorder; proliferative disorder;  
XX KM differential disorder; developmental disorder;  
XX KM haematopoietic disorder;  
XX OS Homo sapiens.  
XX PN US6331427-B1.  
XX PD 18-DEC-2001.

```

PF 26-MAR-1999; 99US-0280116.
PR 26-MAR-1999; 99US-0280116.
XX (MILL-) MILLENNIUM PHARM INC.
XX Robison KE;
XX MPI; 2002-129545/17.
DR New polynucleotides encoding protease homologs of the G-protein-coupled
PT protease family, useful in identifying agonists and antagonists for
PT diagnosis and treatment of protease mediated disorders -
PS Disclosure; Column 109-112; 246pp; English.
XX The invention relates to an isolated human protease nucleic acid molecule
CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
CC defined in the specification. Also disclosed are production of an
CC isolated polypeptide encoded by the nucleic acid, comprising introducing
CC the nucleic acid into a host cell and culturing under conditions to
CC express the protein from the nucleic acid, use of an antibody to
CC detect the encoded protein in a sample and to modulate its in vivo
CC activity, identifying agents that bind to the protein and identification
CC of a polynucleotide agent that modulates the expression of the nucleic
CC acid or its complement (i.e. gene therapy). The nucleic acid can be used
CC to identify an agent that modulates the expression or activity of the
CC nucleic acid, and can be used to isolate the protein. The nucleic acid
CC can be used in diagnostic assays for determining nucleic acid expression
CC as well as activity in the context of a biological sample (e.g., blood,
CC serum, cells, tissue) to determine whether an individual has a disease or
CC disorder, or is at risk of developing a disease or disorder, associated
CC with aberrant expression or activity of the nucleic acid. The nucleic
CC acid can be used to detect mutations in protease genes and gene
CC expression products such as mRNA. The nucleic acid can be used as
CC hybridisation probes to detect naturally-occurring genetic mutations in
CC a protease gene. The nucleic acid can be used in drug screening methods
CC to identify agonists and antagonists that can be used to diagnose and
CC treat such protease mediated disorders e.g., proliferative, the nucleic
CC differential, developmental or haematopoietic disorders. The nucleic
CC acid can be used as probes, primers, in biological assays, to determine
CC patterns of gene expression, to design ribozymes and to construct
CC transgenic animals. The present sequence represents one of the 268
CC disclosed human G-protein-coupled protease cDNA sequences.
XX Sequence 865 BP; 207 A; 197 C; 225 G; 236 T; 0 other;
SQ

```

Query Match 96.3%; Score 834; DB 24; Length 865;  
Best Local Similarity 98.7%; Pred. No. 1.7e-233;  
Matches 851; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```

Qy 1 GGCACCTACTCCCTGAGCTAAGGGGGAAGAGTGCATCCATGAATATGTTCTTAT 60
Db 865 GGCACCTACTCCCTGAGCTAAGGGGGAAGAGTGCATCCATGAATATGTTCTTAT 806
Qy 61 TGGGTGTCTCTGCTGGGACATTTTCTTGTGCTGATCATCTGTTGAGAAAGAACCTG 120
Db 805 TGGGTGTCTCTGCTGGGACATTTTCTTGTGCTGATCATCTGTTGAGAAAGAACCTG 746
Qy 121 CTCCTAATTTGTTGATCACTCAAGTCTCACTTCAACCCCTGTGTGGGCGTCTCATCAAC 180
Db 745 CTCCTAATTTGTTGATCACTCAAGTCTCACTTCAACCCCTGTGTGGGCGTCTCATCAAC 686
Qy 181 CCAGCTGGGTGTGGCCCCAGCTCACTGTATTATTAACCAATCTGAAGTATGCTGGAA 240
Db 685 CCAGCTGGGTGTGGCCCCAGCTCACTGTATTATTAACCAATCTGAAGTATGCTGGAA 626
Qy 241 ATTTCAAGACAGATCAAGACGTAAGTCAAGACAAATTAACCCATTCAAGTCTCC 300
Db 625 ATTTCAAGACAGATCAAGACGTAAGTCAAGACAAATTAACCCATTCAAGTCTCC 566
Qy 301 GCTACTGGAATCACTCATAGAGGCCCCACAGATGACTCTATGCTCATCAAGCTGGCTA 360

```



Db 565 GCTACTGGAAGTACAGTCATAGCCGCCCAAGATGACCTCATCATCAAGCTGCTA 506  
 Qy 361 AGCTGCGCATGCTCATATCCCAAGTCCAGCCCTT-CGCTGCGCACCAACATGTCAG 419  
 Db 505 AGCTGCGCATGCTCATATCCCAAGTCCAGCCCTTACCCTTCCGCAACCAAGTTCAG 446  
 Qy 420 CCAGGCACTGCTGCTCTACTCTCAAGTTTGAGCTGAGCCCAAGAAAACAGTGCAGAC 479  
 Db 445 CCGGCACTGCTGCTCTACTCTCAAGTTTGAGCTGAGCCCAAGAAAACAGTGCAGAC 386  
 Qy 480 CCTGACTTGGGCGGAGAACCTTGAAGGCCCGGTGATGCTGATGAGAAATGCAAAAAA 539  
 Db 385 CTTACTTGGCGGAGAACCTTGAAGGCCCGGTGATGCTGATGAGAAATGCAAAAAA 326  
 Qy 540 GAACAAGAAAAGAACCAAGAAATCTTATGTTGAATTTGAAAGTATTCAGCCGA 599  
 Db 325 GAACAAGAAAAGAACCAAGAAATCTTATGTTGAATTTGAAAGTATTCAGCCGA 266  
 Qy 600 ATTTTGGGAGGTGGCCGTTGCTACTGTCATCTGCAAAAGCAAGCTCCAGGAAATGAG 659  
 Db 265 ATTTTGGGAGGTGGCCGTTGCTACTGTCATCTGCAAAAGCAAGCTCCAGGAAATGAG 206  
 Qy 660 GTGGGCACTTCATGAGGAGGAGGAGCTGGGCAATCCATGTTTACAAATATGATCC 719  
 Db 205 GTGGGCACTTCATGAGGAGGAGGAGCTGGGCAATCCATGTTTACAAATATGATCC 146  
 Qy 720 TGAATTGAGAACTGCTTAAGAGCAAGTGAAGCCCTACTCTCCCTGTCATTCACATG 779  
 Db 145 TGAATTGAGAACTGCTTAAGAGCAAGTGAAGCCCTACTCTCCCTGTCATTCACATG 86  
 Qy 780 CTCTGCAATGAGCTATACAGCAAGATATTTCCCTCTATTCAAAATTAATCTCCAAAT 839  
 Db 85 CTCTGCAATGAGCTATACAGCAAGATATTTCCCTCTATTCAAAATTAATCTCCAAAT 26  
 Qy 840 GAAAATTGGAGTATGACATA 861  
 Db 25 GAAAATTGGAGTATGACATA 4  
 RESULT 8  
 ID ABA04594 standard; cDNA; 730 BP.  
 XX ABA04594:  
 AC ABA04594:  
 XX 21-FEB-2002 (first entry)  
 DT XX  
 DE MOL6a coding sequence.  
 XX  
 KW MOL; G-coupled protein-receptor; cardiomyopathy; atherosclerosis;  
 KW cell signal processing; metabolic disorder; diabetes; cancer;  
 KW neurodegenerative disorder; immune disorder; cardiac disorder;  
 KW lung disease; autoimmune disease; developmental disorder; anti-diabetic;  
 KW Cytoskeletal; Neuroprotective; Antiatherosclerotic; Immunosuppressive;  
 KW Gene therapy; Vaccine; antiinflammatory; MOL6a; trypsin-like; ss.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT 8..715  
 FT /\*tag= a  
 FT /product= "MOL6a"  
 PN MO200181578-A2.  
 PD 01-NOV-2001.  
 XX  
 PF 26-APR-2001; 2001WO-US13578.  
 XX  
 PR 26-APR-2000; 2000US-200158P.  
 PR 28-APR-2000; 2000US-200613P.  
 PR 28-APR-2000; 2000US-200780P.  
 PR 01-MAY-2000; 2000US-201066P.

PR 01-MAY-2000; 2000US-201007P.  
 PR 01-MAY-2000; 2000US-201236P.  
 PR 01-MAY-2000; 2000US-201238P.  
 PR 02-MAY-2000; 2000US-201186P.  
 PR 03-MAY-2000; 2000US-201474P.  
 PR 03-MAY-2000; 2000US-201508P.  
 PR 25-JUL-2000; 2000US-220591P.  
 PR 15-SEP-2000; 2000US-232678P.  
 PR 22-JAN-2001; 2001US-263217P.  
 PR 30-JAN-2001; 2001US-265160P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Vernet CM, Fernandes ER, Gerlach V, Shinkets RA, Malpankar UM,  
 PI Boldo FL, Zernusen BD, Spytek KA, Majumder K, Tchernev VT;  
 PI Padiguru M, Patturajan M, Burgess CE, Gangolli EA, Smlthson G;  
 PI Rastelli L, Macdougall JR, Taupier RJ, Grose WM, Szekeres ES;  
 PI Alsobrook JP;  
 XX  
 DR WPI; 2002-049278/06.  
 DR P-PSDB; AAM47664.  
 XX  
 PT Novel G-protein coupled receptor-related polypeptides and  
 PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,  
 PT atherosclerosis, disorders related to cell signal processing and for  
 PT identifying modulators -  
 XX  
 PS Claim 8; Page 43-44; 227pp; English.  
 XX  
 CC The present invention relates to novel G-coupled protein-receptor related  
 CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,  
 CC ABA04589-ABA04603 and AAM47659-AAM47673). MOLX proteins and coding  
 CC sequences are useful for treating or preventing a MOLX-associated  
 CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to  
 CC cell signal processing and metabolic pathway modulation, diabetes and  
 CC cancer. Additionally, MOLX proteins and coding sequences are useful for  
 CC preventing and treating a variety of disorders including metabolic  
 CC disorders, nutritional oedema, chronic and hereditary pancreatitis,  
 CC obesity, infectious disease, anorexia, neurodegenerative disorders,  
 CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,  
 CC hematopoietic disorders and various dyslipidaemias, metabolic syndrome X  
 CC and wasting disorders associated with chronic diseases and cancers,  
 CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiodysplasia  
 CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,  
 CC multiple sclerosis, lung diseases including asthma, Crohn's disease,  
 CC scleroderma, autoimmune diseases, developmental disorders and neural tube  
 CC defects. The present sequence is the coding sequence for MOL6a.  
 CC  
 CC MOL6a is a trypsin-like protein.  
 XX  
 SQ Sequence 730 BP; 191 A; 197 C; 170 G; 172 T; 0 other;  
 Query Match 84.3%; Score 730; DB 24; Length 730;  
 Best Local Similarity 100.0%; Pred. No. 4,1e-203; Indels 0; Gaps 0;  
 Matches 730; Conservative 0; Mismatches 0;  
 Qy 35 GATCACCATTGAATATGCTCTTATTTGGGTGCTCTGCGACATTTTCTTCTGA 94  
 Db 1 GATCACCATTGAATATGCTCTTATTTGGGTGCTCTGCGACATTTTCTTCTGA 60  
 Qy 95 CTCATCTGTTGAGAAAGAACCCCTGCTCCCTATTTGGTGTACTGAAGTCTCACTTCA 154  
 Db 61 CTCATCTGTTGAGAAAGAACCCCTGCTCCCTATTTGGTGTACTGAAGTCTCACTTCA 120  
 Qy 155 CCCCTGTGGGGGCTCTCATCAAAACCGAGTGGGTGCTGCCCCGCTCACTGCTATTT 214  
 Db 121 CCCCTGTGGGGGCTCTCATCAAAACCGAGTGGGTGCTGCCCCGCTCACTGCTATTT 180  
 Qy 215 ACCAATCTGAAGTATGCTGGGAATTTCAAGACAGAGTCAAGAGCGTACTGAACA 274  
 Db 181 ACCAATCTGAAGTATGCTGGGAATTTCAAGACAGAGTCAAGAGCGTACTGAACA 240  
 Qy 275 GACAATTAACCCCATTCAGATGCTCGCTACTGAACTACAGTCATAGCGCCACAGGA 334

```

Db      241 GACAAATTAACCCCATTCAGATTCCTCGCTACTGGAACCTACATCATAGCGCCACAGAGA 300
Qy      335 TGACCTCATGCTCATAGCTGGCTAAGCTGSCATGCTCAATCCAAATGCCAGCCCT 394
Db      301 TGACCTCATGCTCATAGCTGGCTAAGCTGSCATGCTCAATCCAAATGCCAGCCCT 360
Qy      395 TCCCTCGCCACCAACCAATGTCAGGCGAGCACTGCTGCTACTACTCAAGTTGGACTG 454
Db      361 TCCCTCGCCACCAACCAATGTCAGGCGAGCACTGCTGCTACTACTCAAGTTGGACTG 420
Qy      455 GAGCCAAAGAAACAGTGGCCGACACCTCTGAGCGAGAACCTGAGGCCCCCTGAT 514
Db      421 GAGCCAAAGAAACAGTGGCCGACACCTCTGAGCGAGAACCTGAGGCCCCCTGAT 480
Qy      515 GTCGTGATCGAGATGCGCAAAAACAGAAACAGAGAAAAGCCAGAAATTCCTTATGCT 574
Db      481 GTCGTGATCGAGATGCGCAAAAACAGAAACAGAGAAAAGCCAGAAATTCCTTATGCT 540
Qy      575 GAAATTTGAAAGTATTCAGCCGAATTTTGGGAGTGGCCGTTGCTACTGTCATCTG 634
Db      541 GAAATTTGAAAGTATTCAGCCGAATTTTGGGAGTGGCCGTTGCTACTGTCATCTG 600
Qy      635 CAAAGCAAGCTCCAGGAAATCGAGGTGGGCACTTCATGAGAGGAGCGCTGGCATCTA 694
Db      601 CAAAGCAAGCTCCAGGAAATCGAGGTGGGCACTTCATGAGAGGAGCGCTGGCATCTA 660
Qy      695 CACCAATGTTTAAATATCTATCTGATGAGAAACATGCTGTAAGGACAAAGTGAACCC 754
Db      661 CACCAATGTTTAAATATCTATCTGATGAGAAACATGCTGTAAGGACAAAGTGAACCC 720
Qy      755 TACTTCTCCC 764
Db      721 TACTTCTCCC 730

```

RESULT 9  
ABAO4595  
ID ABAO4595 standard; cDNA; 730 BP.  
XX ABAO4595;  
AC ABAO4595;  
XX ABAO4595;  
DT 21-FEB-2002 (first entry)  
XX  
DE MOLdb coding sequence.  
XX  
XX MOL: G-coupled protein-receptor; cardiomyopathy; atherosclerosis;  
KM cell signal processing; metabolic disorder; diabetes; cancer;  
KM neurodegenerative disorder; immune disorder; cardiac disorder;  
KM lung disease; autoimmune disease; developmental disorder; anti-diabetic;  
KM Cystoelastic; Neuroprotective; Antiatherosclerotic; Immunosuppressive;  
KM Gene therapy; Vaccine; anti-inflammatory; MOLdb; trypsin-like; ss.  
XX  
XX Unidentified.  
OS  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 8..715  
FT /tag= a  
FT /product= "MOLdb"  
XX  
XX  
PN MO200181578-A2.  
PD 01-NOV-2001.  
XX  
XX 26-APR-2001; 2001WO-US13578.  
XX  
XX 26-APR-2000; 2000US-200158P.  
PR 28-APR-2000; 2000US-200613P.  
PR 28-APR-2000; 2000US-200780P.  
PR 01-MAY-2000; 2000US-201006P.  
PR 01-MAY-2000; 2000US-201007P.  
PR 01-MAY-2000; 2000US-201236P.  
PR 01-MAY-2000; 2000US-201238P.  
PR 02-MAY-2000; 2000US-201186P.

```

PR      03-MAY-2000; 2000US-201474P.  
PR      03-MAY-2000; 2000US-201508P.  
PR      25-JUL-2000; 2000US-220591P.  
PR      15-SEP-2000; 2000US-232678P.  
PR      22-JAN-2001; 2001US-263217P.  
PR      30-JAN-2001; 2001US-265160P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Vernet CM, Fernandes ER, Gerlach V, Shimkete RA, Malpankar UM,  
PI Boldog FL, Zernhosen BD, Spytek KA, Majumder K, Tchernev VT;  
PI Padigaru M, Patlurajan M, Burgess CE, Gangoli EA, Smithson G;  
PI Raselli L, Macdougall JR, Taupier RJ, Grose WM, Szekeres ES;  
PI Alsdbrook JP;  
XX  
XX WPI, 2002-049278/06.  
DR P-PSDB; AAM47665.  
XX  
XX Novel G-protein coupled receptor-related polypeptides and  
PT polynucleotides for diagnosing, preventing and treating cardiomyopathy,  
PT atherosclerosis, disorders related to cell signal processing and for  
PT identifying modulators  
XX  
XX Claim 8; Page 45; 227p; English.  
XX  
XX The present invention relates to novel G-coupled protein-receptor related  
CC proteins and coding sequences (MOLX, where X is a number from 1 to 10,  
CC ABAO4589-ABAO4603 and AAM47659-AAM47673). MOLX proteins and coding  
CC sequences are useful for treating or preventing a MOLX-associated  
CC disorder, such as cardiomyopathy, atherosclerosis, disorders related to  
CC cell signal processing and metabolic pathway modulation, diabetes and  
CC cancer. Additionally, MOLX proteins and coding sequences are useful for  
CC preventing and treating a variety of disorders including metabolic  
CC disorders, nutritional oedema, chronic and hereditary pancreatitis,  
CC obesity, infectious disease, anorexia, neurodegenerative disorders,  
CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,  
CC haematopoietic disorders and various dyslipidaemias, metabolic syndrome X  
CC and wasting disorders associated with chronic diseases and cancers,  
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiodenias  
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,  
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,  
CC scleroderma, autoimmune diseases, developmental disorders and neural tube  
CC defects. The present sequence is the coding sequence for MOLdb.  
XX  
XX MOLdb is a trypsin-like protein.  
XX  
XX Sequence 730 BP; 191 A; 198 C; 170 G; 171 T; 0 other;  
SQ  
XX  
XX Query Match 84.1%; Score 728.4; DB 24; Length 730;  
XX Best Local Similarity 99.9%; Pred. No. 1.2e-202;  
XX Matches 729; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 35 GATCACCATGAATAATGCTTCTATTGGGTGCTCCTCGTGGACATTTTCTTGTCTGA 94  
Db 1 GATCACCATGAATAATGCTTCTATTGGGTGCTCCTCGTGGACATTTTCTTGTCTGA 60  
Qy 95 CTCATCTGTTCAAGAAAGAACCCCTGCTCTCTATTGGTGTACTTCAATCTCACTTCAA 154  
Db 61 CTCATCTGTTCAAGAAAGAACCCCTGCTCTCTATTGGTGTACTTCAATCTCACTTCAA 120  
Qy 155 CCCCTGTGTGGGCGTCTCATCAAAACCCAGCTGGGTGCTGGCCCGCCAGCTCACTGCTATT 214  
Db 121 CCCCTGTGTGGGCGTCTCATCAAAACCCAGCTGGGTGCTGGCCCGCCAGCTCACTGCTATT 180  
Qy 215 ACCCAATCTGAAGTATGCTGGGAAATTTCAAGAGCAGAGTCAAGACGGTACTGAACA 274  
Db 181 ACCCAATCTGAAGTATGCTGGGAAATTTCAAGAGCAGAGTCAAGACGGTACTGAACA 240  
Qy 275 GACCAATTAACCCCATTCAGATGTCGCTACTGGAACCTACATGATCAAGGCCCCCAGAGA 334  
Db 241 GACCAATTAACCCCATTCAGATGTCGCTACTGGAACCTACATGATCAAGGCCCCCAGAGA 300  
Qy 335 TGACCTCATGCTCATAGCTGGCTAAGCTGSCATGCTCAATCCAAATGCCAGCCCT 394

```

Db 301 TGACCTCATGCTCATCAAGCTGGCTAAGCCTGACCTCAATCCAAAGTCCAGCCCT 360  
Qy 395 TCCCTGCGCCACCAATATGTCAGGCGAGCACTGTCTGTACTCTCAGGTTTGAATG 454  
Db 361 TCCCTGCGCCACCAATATGTCAGGCGAGCACTGTCTGTACTCTCAGGTTTGAATG 420  
Qy 455 GAGCGCAAGAAAAGTGGCGGACACCTGACTTGGCGCAAGACCTGGAGGCCCGCGTAT 514  
Db 421 GAGCGCAAGAAAAGTGGCGGACACCTGACTTGGCGCAAGACCTGGAGGCCCGCGTAT 480  
Qy 515 GTCTGATCGGAAGTCCCAAAAAACAGAACAGAAAAGCCACAGAAATCTTATATGT 574  
Db 481 GTCTGATCGGAAGTCCCAAAAAACAGAACAGAAAAGCCACAGAAATCTTATATGT 540  
Qy 575 GAAATTTGTGAAGTATTTAGCCGCAATTTTGGGAGAGTGGCCCTTCTACTGTCACTG 634  
Db 541 GAAATTTGTGAAGTATTTAGCCGCAATTTTGGGAGAGTGGCCCTTCTACTGTCACTG 600  
Qy 635 CAAAGACAGCTCCAGGGAATTCAGGTGGGCACTTCATGGAGGAGGACGTCCGATCTA 694  
Db 601 CAAAGACAGCTCCAGGGAATTCAGGTGGGCACTTCATGGAGGAGGACGTCCGATCTA 660  
Qy 695 CACCAATGTTACAAATATGATCTGTGATTTAGAGAACATCTTAAGACAAAGTAGACC 754  
Db 661 CACCAATGTTACAAATATGATCTGTGATTTAGAGAACATCTTAAGACAAAGTAGACC 720  
Qy 755 TACTTCTCCC 764  
Db 721 TACTTCTCCC 730

RESULT 10  
AAS44567  
ID AAS44567 standard; cDNA; 708 BP.

AC AAS44567;

DT 18-DEC-2001 (first entry)

DE Human cDNA encoding protein AFP80526.

KM Human; sex: AFR; cytostatic; Marfan's syndrome; thrombocytopenia;  
leukemia; porphyria; Gilles De la Tourette's syndrome; gene therapy;  
squamous cell carcinoma; diabetes mellitus; Grave's disease;  
colon cancer; Alzheimer's disease; epiphyseal dysplasia.

OS Homo sapiens.

PN MO20016748-A2.

PD 13-SEP-2001.

PF 05-MAR-2001; 2001MO-US07192.

PR 03-MAR-2000; 2000US-187221P.

PA (ZYMO ) ZYMOGENETICS INC.

PI Conklin DC, Presnell SR, Adler DA;

DR WPI; 2001-589943/66.

DR P-PSDB; AAU27666.

PT Novel AFP polypeptides and polynucleotides, useful for diagnostic and  
therapeutic purposes, in cancer therapy and for screening modulator

PT compounds -

PS Claim 10; Page 192-193; 220pp; English.

CC The invention relates to novel human AFP proteins (not defined) and  
CC the nucleic acids that encode them. AFP proteins are useful as standards  
CC in assays of protein and protein inhibitors in both clinical and research  
CC settings, as protein and amino acid supplements, including hydrolysates.

CC The nucleic acids are useful for radiation hybrid mapping. Secretory  
CC fusion proteins of AFP are useful in cancer therapy, for enhancing in  
CC vitro cytotoxicity, for enhancing in vivo killing of target tissues, and  
CC for targeted cells or tissue inhibition or ablation. Anti-AFP antibodies  
CC are useful for isolating target polypeptides by affinity purification, in  
CC diagnostic assays for determining circulating or localised levels of  
CC target polypeptides, for tissue typing, for cell sorting, for screening  
CC expression libraries, for generating anti-idiotypic antibodies, and as  
CC neutralising antibodies or as antagonists to block protein activity in  
CC vitro and in vivo. AFP proteins and nucleic acids may be used to  
CC diagnose or treat (e.g. by gene therapy) diseases associated with the  
CC malfunction of the AFP e.g. AFP16824 and Marfan's syndrome,  
CC AFP576853/AFP3158 and thrombocytopenia, leukemia, porphyria, Gilles De  
CC La Tourette's syndrome, AFP552829 and squamous cell carcinoma, diabetes  
CC mellitus, Grave's disease, AFP64311 and colon cancer, AFP253034 and  
CC Alzheimer's disease and AFP66580 and epiphyseal dysplasia. Many more  
CC examples of disease are given in the specification. The present  
CC sequence encodes an AFP of the invention.

XX Sequence 708 BP; 188 A; 185 C; 168 G; 167 T; 0 other;

Query Match 81.6%; Score 706.4; DB 22; Length 708;  
Best Local Similarity 99.9%; Pred. No. 3.2e-196;  
Matches 707; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 42 ATGAATATGCTCTTCTATTGGGTGTCCTCGTGGAGCAATTTCTTTCGTGACTCATCT 101  
Db 1 ATGAATATGCTCTTCTATTGGGTGTCCTCGTGGAGCAATTTCTTTCGTGACTCATCT 60  
Qy 102 GTTCAGAAAAGAACCCCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT 161  
Db 61 GTTCAGAAAAGAACCCCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT 120  
Qy 162 GTGGGGGTCTCTATCAACCCAGCTGGGTGGTGGCCCGCCAGCTCACTGATTTACCAAT 221  
Db 121 GTGGGGGTCTCTATCAACCCAGCTGGGTGGTGGCCCGCCAGCTCACTGATTTACCAAT 180  
Qy 222 CTGAAGATGATGCTGGGAATTTCAAGACAGAGTCAAGAGCGTACTGAACAGCAAT 281  
Db 181 CTGAAGATGATGCTGGGAATTTCAAGACAGAGTCAAGAGCGTACTGAACAGCAAT 240  
Qy 282 AACCCATTCAGATGCTCCGCTACTGTGAACATCAAGTCAATAGGCCCCACAGATGACCTC 341  
Db 241 AACCCATTCAGATGCTCCGCTACTGTGAACATCAAGTCAATAGGCCCCACAGATGACCTC 300  
Qy 342 ATGCTATCAAGTGGCTTAAGCCTGCAAGTCTCAATCCCAAGTCCAGGCCCTTCCCTC 401  
Db 301 ATGCTATCAAGTGGCTTAAGCCTGCAAGTCTCAATCCCAAGTCCAGGCCCTTCCCTC 360  
Qy 402 GCCACCAAGATGTCAGGCGAGCACTGTCTGTCTACTCTCAGGTTTGGAGTGAACCAA 461  
Db 361 GCCACCAAGATGTCAGGCGAGCACTGTCTGTCTACTCTCAGGTTTGGAGTGAACCAA 420  
Qy 462 GAAACAGTGGCGGACACCTGACTTGGCGGAGAACCTGGAAGGCCCGCTGATGTGAT 521  
Db 421 GAAACAGTGGCGGACACCTGACTTGGCGGAGAACCTGGAAGGCCCGCTGATGTGAT 480  
Qy 522 CGAATGCGCAAAAACAGAACAGAAAAGCCACAGAAATCTTATGTGAAATT 581  
Db 481 CGAATGCGCAAAAACAGAACAGAAAAGCCACAGAAATCTTATGTGAAATT 540  
Qy 582 GTGAAGTATTCAGCGGAATTTTGGGAGTGGCCGTTGCTACTGTCACTGCAAGAC 641  
Db 541 GTGAAGTATTCAGCGGAATTTTGGGAGTGGCCGTTGCTACTGTCACTGCAAGAC 600  
Qy 642 AAGCTCAGGGAATGAGGTGGGCACTTCATGGAAGGAGACGTGGCATTTACCAAT 701  
Db 601 AAGCTCAGGGAATGAGGTGGGCACTTCATGGAAGGAGACGTGGCATTTACCAAT 660  
Qy 702 GTTTACAAATATGATCTCTGATTTGAACATCTGCTAAGACAAAGTGA 749  
Db 661 GTTTACAAATATGATCTCTGATTTGAACATCTGCTAAGACAAAGTGA 708



XX 23-APR-2002 (first entry)  
 DT DNA encoding novel human protease #43.  
 XX  
 DE  
 XX Human; protease; cancer; immune-related disorder; cardiovascular disease;  
 KM neuronal-associated disease; metabolic disorder; inflammatory disorder;  
 KM nervous system disorder; sexual dysfunction; pain; mood disorder;  
 KM hypertension; psychotic disorder; neurological disorder; dyskinesia;  
 KM viral infection; human immunodeficiency virus; HIV; non-viral infection;  
 KM ocular disease; cytoskeletal; gene; de.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200200860-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 26-JUN-2001; 2001MO-US20171.  
 XX  
 PR 26-JUN-2000; 2000US-214047P.  
 XX  
 PA (SUGF-) SUGEN INC.  
 XX  
 PI Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
 PI Charyczak G;  
 XX  
 DR P-PSDB; AAU82744.  
 XX  
 XX MPI: 2002-139913/18.  
 XX  
 PT Nucleic acids encoding novel human proteases, useful for useful for  
 PT treating diseases and disorders such as cancers, immune-related  
 PT diseases and disorders, cardiovascular disease (e.g. restenosis) and  
 PT inflammatory disorders -  
 XX  
 PS Claim 26; Fig 1MM; 313pp; English.  
 XX  
 XX The present invention relates to the isolation of novel human  
 CC proteases, and the nucleic acids encoding them. The sequences of  
 CC the invention are useful for treating diseases and disorders such as  
 CC cancers (e.g. breast, colon, lung), immune-related diseases and disorders  
 CC (e.g. inflammatory diseases and asthma), cardiovascular diseases  
 CC (e.g. restenosis and coronary thrombosis), brain or neuronal-associated  
 CC diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory  
 CC disorders (e.g. rheumatoid arthritis and psoriasis), central or  
 CC peripheral nervous system diseases, migraines, pain, sexual dysfunction,  
 CC mood disorders, attention disorders, cognition disorders, hypotension,  
 CC hypertension, psychotic disorders, neurological disorders and dyskinesias.  
 CC (e.g. Alzheimer's disease, Parkinson's disease) and dyskinetic.  
 CC The nucleic acids and polypeptides are also useful for treating viral  
 CC infections caused by human immunodeficiency virus (HIV), and non-viral  
 CC infections such as ocular disease (e.g. glaucoma) and macular  
 CC degeneration. ABK31744-ABK31802 represent DNA sequences encoding for  
 CC the novel human proteases of the invention.  
 CC  
 XX Sequence 798 BP; 212 A; 212 C; 191 G; 183 T; 0 other;  
 XX  
 SQ  
 Query Match 70.0%; Score 606.4; DB 24; Length 798;  
 Best Local Similarity 88.6%; Pred. No. 6.1e-167;  
 Matches 707; Conservative 0; Mismatches 1; Indels 90; Gaps 1;  
 QY 42 ATGAATATGCTCTTATTTGGGTGCTCGTGGAGACATTTTCTTGGTGAATCATCT 101  
 DB 1 ATGAATATGCTCTTATTTGGGTGCTCGTGGAGACATTTTCTTGGTGAATCATCT 60  
 QY 102 GTTCAGAAAGAACCTGCTCCCTATTTGGTGTACTCAAGTCACTTCAACCCCTGT 161  
 DB 61 GTTCAGAAAGAACCTGCTCCCTATTTGGTGTACTCAAGTCACTTCAACCCCTGT 120  
 QY 162 GTGGGGTCTCTATCAAACTGGGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 221  
 DB 121 GTGGGGTCTCTATCAAACTGGGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 180

QY 222 CTGAAGATGATGCTGGAAATTTTCAAGACAGAGTCAGAGACGGTACTGAACAGCAATT 281  
 DB 181 CTGAAGATGATGCTGGAAATTTTCAAGACAGAGTCAGAGACGGTACTGAACAGCAATT 240  
 QY 282 AACCCATTCAGATTCGCTCCGCTACTGAGAACTAAGTCATAGCCGCCACAGATGACCTC 341  
 DB 241 AACCCATTCAGATTCGCTCCGCTACTGAGAACTAAGTCATAGCCGCCACAGATGACCTC 300  
 QY 342 ATGCTCATCAAGCTGAGCTAAGCTGACATGCTCAATCCCAAAATGCAAGCCCTTCCCTC 401  
 DB 301 ATGCTCATCAAGCTGAGCTAAGCTGACATGCTCAATCCCAAAATGCAAGCCCTTCCCTC 360  
 QY 402 GCCACCAACCAATGTCAGGCCAGGCACTGCTGTCTACTCTGAGTTTGGACTGAGCCAA 461  
 DB 361 GCCACCAACCAATGTCAGGCCAGGCACTGCTGTCTACTCTGAGTTTGGACTGAGCCAA 420  
 QY 462 GAAAACAGT-----GGCCGACACCCTTGCCG 491  
 DB 421 GAAAACAGTGGGCTTTGGCACTGAGCCAGGCCATCTGACTCTGCAAGAGGCCCA 480  
 QY 471 -----GGCCGACACCCTTGCCG 491  
 DB 481 GCCATTCCTGATTTGCGAGACACAAATTCAGATGAACAAAGCCGACACCTTGACTTCCG 540  
 QY 492 CAGAACCTGAGAGCCCGCTGATGTGATCGAATGCCCCAAAACAGAACAGAGAAA 551  
 DB 541 CAGAACCTGAGAGCCCGCTGATGTGATCGAATGCCCCAAAACAGAACAGAGAAA 600  
 QY 552 AGCCACAGGAATTCCTTATGTGAAATTTTGGAAATTCAGCCGAAATTTTGGGGAG 611  
 DB 601 AGCCACAGGAATTCCTTATGTGAAATTTTGGAAATTCAGCCGAAATTTTGGGGAG 660  
 QY 612 GTGGCGGTGCTACTGTCATCTGCAAAAGCAAGCTCCAGGAAATCGAGTGGGGCACTTC 671  
 DB 661 GTGGCGGTGCTACTGTCATCTGCAAAAGCAAGCTCCAGGAAATCGAGTGGGGCACTTC 720  
 QY 672 ATGGAGGGGAGCGTGGCATTTACACCAATGTTTCAAAATATGTATCTTGATTTGAAAC 731  
 DB 721 ATGGAGGGGAGCGTGGCATTTACACCAATGTTTCAAAATATGTATCTTGATTTGAAAC 780  
 QY 732 ACTGCTAAGACAAAGTGA 749  
 DB 781 ACTGCTAAGACAAAGTGA 798  
 RESULT 13  
 AAS12971  
 ID AAS12971 standard; DNA; 705 BP.  
 XX  
 AC AAS12971;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human Zt1p3 serine protease degenerate DNA.  
 XX  
 KM Human; Zt1p3; serine protease; asthma; vascular function; inflammation;  
 KM gene therapy; stroke; testicular function; spermatogenesis; haemostatic;  
 KM mass spectrometry; circular dichroism; X-ray crystallography; ds;  
 KM nuclear magnetic resonance spectroscopy; antiasthmatic; anti-inflammatory;  
 KM cerebroprotective.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 PN W0200166771-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 28-FEB-2001; 2001MO-US06432.  
 XX  
 PR 03-MAR-2000; 2000US-0518387.  
 XX  
 PA (ZYMO ) ZYMOGENETICS INC.

XX Conklin DC;  
 PI MPI; 2001-589946/66.  
 XX  
 DR  
 XX  
 PT Novel Ztryp3 polypeptides and polynucleotides useful in the treatment  
 PT of asthma, vascular disorders including stroke, inflammation and  
 PT testicular function  
 PS Disclosure; Page 82; 82pp; English.  
 XX  
 CC The invention relates to an isolated human Ztryp3 polypeptide, a member  
 CC of the serine protease family. Ztryp3 polypeptides and their associated  
 CC polynucleotides are useful in diagnosis, therapy and industry and are  
 CC used as targets for identifying modulators, preferably inhibitors of  
 CC serine protease activity. The sequences are useful in the treatment of  
 CC asthma, vascular function such as stroke, inflammation and testicular  
 CC function (by modulating spermatogenesis). Ztryp3 proteins can be used for  
 CC identifying peptide cleavage sites and for coupling amino and carboxy  
 CC terminal tags. The polypeptides are also useful to teach analytical  
 CC skills such as mass spectrometry, circular dichroism, X-ray  
 CC crystallography and nuclear magnetic resonance spectroscopy. This  
 CC sequence represents degenerate DNA encoding human Ztryp3.  
 XX  
 SQ Sequence 705 BP; 133 A; 68 C; 116 G; 102 T; 286 other;  
 Query Match 59.1%; Score 511.4; DB 22; Length 705;  
 Best Local Similarity 59.3%; Pred. No. 3.6e-139;  
 Matches 418; Conservative 164; Mismatches 123; Indels 0; Gaps 0;  
 Oy 42 ATGAAATATGCTCTTCTGAGTGGTCTGCTGCGAGACATTTTCTGCTGACATCATC 101  
 Db 1 ATGAARTATGTTTATYATYNGNGNTYNGCNGAATTTTCTGCGAAYNSMNSN 60  
 Oy 102 GTTCAGAAAGAACCCCTGCTCTCTATTTGGTGTACCTCAAGTCTCACTTCAACCCCTGT 161  
 Db 61 GTNCARAAGARGAYCCNGCNCNTATYNTGTATYTAARWSNCAATTTAAAYCCTGT 120  
 Oy 162 GTGGGGGCTCTCAACCCAGCTGGGCTGGCCCGCCAGCTCACTGCTATTTACCAAT 221  
 Db 121 GTNGGNGTATTAATTAARCCNSNTGGTYNTGNCNGCNCATYGTATYNNCCNAAY 180  
 Oy 222 CTGAAGATGATGCTGAGCAATTTCAAGAGAGAGAGACGCTAGTCAAGACAGCAAT 281  
 Db 181 YTNAAAGTATGTTGNSNAATTTAAARMSMGNTMGNGAYGNAACNARCARACNAH 240  
 Oy 282 AACCCCATTCAGATGCTCGCTACTGGAATCAAGTACATAGCCGCCACAGATGACCTC 341  
 Db 241 AAYCNAATCARATHGTNNMNTATYTGAAATYATWSNCAVWSNCCNCNCARAYGAYTN 300  
 Oy 342 ATGCTATCAAGTGTCTAAGCTGCAATGCTCAATCCAAAGTCCAGCCCTTCCCTC 401  
 Db 301 ATGYTATTAARATYNNCNAARCCNGCNAATGTAAACNAAARTNARCCNATYNA 360  
 Oy 402 GCACACCAATGCTGAGGACGACATGCTGTCTACTCTCAGGTTGGACGTGAGCCAA 461  
 Db 361 GCNACNAHAAYTNNMNCNGNACNGTGTATYTTTMSKNGANTNGATYTGMSNAR 420  
 Oy 462 GAAACAGTGGCGACACCTGACTTGGCGAGACACCTGAGAGCCGCCGTGATGTGAT 521  
 Db 421 GAAATWSNNGNNGNCAVCCNGAYTNNMGCAAAATYTNGARCCNCNGNTATGMSNGAY 480  
 Oy 522 CGAGATGCGCAAAAAACAGAACAGAAAAAGCCACAGAAATTCCTTATGTGAAATTT 581  
 Db 481 MNGNATGTCARAAACNARCARAGNAAARWSNCAVWSNCAVWSNCAVWSNCAVWSN 540  
 Oy 582 GTGAAGATGTTGAGCGAATTTTGGGAGGTGGCGGCTGTCTGCTGATCTGCAATGCAAGC 641  
 Db 541 GTNAAGTATTTTWSNMGNTHTTGGNGARGTNGCNGTNGCNAACNGTATHTAARAY 600  
 Oy 642 AAGCTTCAGAGATTCAGGTGGGCACTTCATGGAGGGGACCTCGGCACTTCAACCAAT 701  
 Db 601 AARYTNCAGNATHTGARGTNGNCATYTTATGGGNGGAGATYTNNGNATHTTAAACNAAY 660

Oy 702 GTTTCAATATGATCTGATTTGAGAACACTGCTAGACAG 746  
 Db 661 GTTATTAATATGTTMSNTGATHTGARAYACNCAAGATYAR 705  
 RESULT 14  
 ID ABQ77083  
 XX ABQ77083 standard; DNA; 705 BP.  
 AC  
 XX  
 DT ABQ77083;  
 XX  
 DE 01-APR-2003 (first entry)  
 XX  
 DE Human serine protease Ztryp3 coding region.  
 XX  
 KW Human, serine protease; Ztryp3; gene; blood coagulation; fibrinolysis;  
 KW complement activation; fertilization; hormone production; gene therapy;  
 KW somatic cell; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6468776-B1.  
 PD 22-OCT-2002.  
 XX  
 PF 28-FEB-2001; 2001US-0796110.  
 XX  
 PR 03-MAR-2000; 2000US-186623P.  
 XX  
 PA (ZYMO ) ZYMOGENETICS INC.  
 PI Conklin DC;  
 XX  
 DR MPI; 2003-182091/18.  
 DR P-PSDB; ABG73741.  
 XX  
 PT Novel serine protease nucleic acid molecule useful for producing serine  
 PT protease polypeptide, and as educational tool in laboratory practicum  
 PT kits for courses related to genetics and molecular biology  
 PS  
 PS Disclosure; Column 59-60; 32pp; English.  
 XX  
 CC This invention describes a novel human serine protease, Ztryp3. The  
 CC serine protease family play a role in carefully controlled processes such  
 CC as blood coagulation, fibrinolysis, complement activation, fertilization  
 CC and hormone production. The products of the invention are useful for  
 CC producing serine protease polypeptides; as probes or primers to clone 5'  
 CC non-coding regions of Ztryp3 gene; as polymerase chain reaction primers  
 CC and probes; as educational tool in laboratory practicum kits for courses  
 CC related to genetics and molecular biology, protein chemistry, and  
 CC antibody production and analysis; as standards or as unknowns for testing  
 CC purposes; as an aid to teach a student how to prepare expression  
 CC constructs for bacterial, viral, or mammalian expression, including  
 CC fusion constructs, where Ztryp3 is the gene to be expressed; for  
 CC determining the restriction endonuclease cleavage sites of the  
 CC polynucleotides; for determining mRNA and DNA localization of Ztryp3  
 CC polynucleotides in tissues (i.e., by northern and southern blotting as  
 CC well as polymerase chain reaction); and for identifying related  
 CC polynucleotides by nucleic acid hybridization. The products of the  
 CC invention can also be used for determining mutations in the Ztryp3 gene,  
 CC for determining Ztryp3 gene expression, for examining Ztryp3 gene  
 CC structure, for detecting and localizing expression of Ztryp3 gene in a  
 CC biological sample, in vivo diagnosis, in gene therapy, particularly  
 CC in somatic cell gene therapy, and for producing transgenic mice. This  
 CC sequence represents the degenerate coding region of the human serine  
 CC protease Ztryp3 described in the method of the invention.  
 XX  
 SQ Sequence 705 BP; 133 A; 68 C; 116 G; 102 T; 286 other;  
 Query Match 59.0%; Score 511.2; DB 25; Length 705;  
 Best Local Similarity 59.3%; Pred. No. 4.1e-139;  
 Matches 418; Conservative 163; Mismatches 124; Indels 0; Gaps 0;



```

QY 42 ATGAATATGCTCTTATTTGGGTGCTGCTGGAGCATTTTCTTGTGTCATCT 101
DB 1 ATGAATATGCTCTTATTTGGGTGCTGCTGGAGCATTTTCTTGTGTCATCT 60
QY 102 GTTCAGAAAGAACCTGCTGCTTATTTGGTGTACTCAAGTCTCAACCTGCT 161
DB 61 GTTCAGAAAGAACCTGCTGCTTATTTGGTGTACTCAAGTCTCAACCTGCT 120
QY 162 GTGGGCGTCTCTCAAAACCCAGCTGGGTGCTGGCCCACTCACTCTTAACCAAT 221
DB 121 GTGGGCGTCTCTCAAAACCCAGCTGGGTGCTGGCCCACTCACTCTTAACCAAT 180
QY 222 CTGAAGTGTGCTGGGAAATTTCAAGACAGAGTCAAGACGCTACTGAACAGCAAT 281
DB 181 YTAAGTATGATGCTGGGAAATTTCAAGACAGAGTCAAGACGCTACTGAACAGCAAT 240
QY 282 AACCCCATTTAGATGCTGCTGCTACTGGAATCAAGTCAAGTCAAGTCAAGTCA 341
DB 241 AACCCCATTTAGATGCTGCTGCTACTGGAATCAAGTCAAGTCAAGTCAAGTCA 300
QY 342 ATGCTATCAAGTGGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401
DB 301 ATGCTATCAAGTGGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 402 GCCACCAACCAATGCTCAAGGCGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461
DB 361 GCNACNACNAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 462 GAAACAGTGGCGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
DB 421 GAAACAGTGGCGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 522 CGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
DB 481 CGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 582 GTGAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
DB 541 GTGAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 642 AACCTCAAGGAAATCGAGTGGGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAG 701
DB 601 AACCTCAAGGAAATCGAGTGGGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAG 660
QY 702 GTTAAATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
DB 661 GTTAAATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705

```

RESULT 15  
AAC24057  
ID AAC24057 standard, cDNA, 379 BP.

AC AAC24057;  
DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 28132.

KM Human, 5' EST, expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (BEST) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI, 2000-500381/45.  
DR  
XX  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1, SEQ ID 28132, 71bp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNA or polyA+ RNA derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
XX Sequence 379 BP; 102 A, 99 C, 81 G, 91 T, 6 other;

Query Match 28.4%; Score 246; DB 21; Length 379;  
Best Local Similarity 92.6%; Pred. No. 1, 2e-61;  
Matches 277; Conservative 2; Mismatches 18; Indels 2; Gaps 2;

```

QY 46 AATATGCTCTTATTTGGGTGCTGCTGGAGCATTTTCTTGTGTCATCTGCTC 105
DB 63 AATATGCTCTTATTTGGGTGCTGCTGGAGCATTTTCTTGTGTCATCTGCTC 122
QY 106 AG-AAAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 164
DB 123 AGAAAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182
QY 165 GGGCTCTCATCAACCCAGCTGGGTGCTGGCCCACTCACTGCTATTTTCAAAATCTG 224
DB 183 GGGCTCTCATCAACCCAGCTGGGTGCTGGCCCACTCACTGCTATTTTCAAAATCTG 242
QY 225 AAGTATGCTGGGAAATTTCAAGACAGAGTCAAGAGAGTCAAGAGAGTCAAGAG 284
DB 243 AAGTATGCTGGGAAATTTCAAGACAGAGTCAAGAGAGTCAAGAGAGTCAAGAG 302
QY 285 CCCATTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 342
DB 303 CCCATTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361

```

Search completed: November 8, 2003, 00:18:32  
Job time : 332 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 21:51:35 ; Search time 78 Seconds  
(without alignments)  
4900.485 Million cell updates/sec

Title: US-10-037-270-482

Perfect score: 866  
Sequence: 1 ggcacttactccctgagctca.....tgggaatgagcatactagt 866

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A COMB. seq: \*  
2: /cgn2\_6/prodata/2/ina/5B COMB. seq: \*  
3: /cgn2\_6/prodata/2/ina/6A COMB. seq: \*  
4: /cgn2\_6/prodata/2/ina/6B COMB. seq: \*  
5: /cgn2\_6/prodata/2/ina/PCTUS COMB. seq: \*  
6: /cgn2\_6/prodata/2/ina/backfiles1. seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	866	100.0	866	4	US-09-620-312D-482
2	850.8	98.2	865	4	US-09-796-110-1
3	834	96.3	865	4	US-09-280-116-63
4	511.2	59.0	705	4	US-09-796-110-3
5	145	16.7	825	3	US-09-120-582-1
6	139.6	16.1	897	2	US-08-956-267A-1
7	66	7.6	1292	4	US-09-205-258-189
8	65.6	7.6	1146	4	US-09-205-258-247
9	65	7.5	452	4	US-09-280-116-64
10	64.8	7.5	1052	4	US-09-386-642-10
11	64.8	7.5	1166	3	US-08-944-483-7
12	64.8	7.5	1192	3	US-08-944-483-8
13	64.8	7.5	1314	3	US-09-025-059-2
14	63.4	7.3	833	2	US-08-790-137-2
15	63	7.3	732	1	US-08-361-395-2
16	60.6	7.0	725	3	US-09-197-801-12
17	60.6	7.0	725	3	US-09-551-028-12
18	55.2	6.4	969	3	US-09-502-600-30
19	55.2	6.4	986	2	US-08-557-146-1
20	55.2	6.4	986	2	US-09-154-344-1
21	55.2	6.4	1089	3	US-08-930-188-1
22	55.2	6.4	1089	3	US-08-930-188-3
23	55.2	6.4	1089	5	PCT-US96-04294-1
24	55.2	6.4	1089	5	PCT-US96-04294-3
25	54	6.2	1386	2	US-08-756-506-3
26	54	6.2	1387	6	5270178-1
27	54	6.2	1755	6	5225537-1

28	49.6	5.7	681	1	US-07-929-198-1	Sequence 1, Appl1
29	49.6	5.7	681	1	US-07-929-198-3	Sequence 3, Appl1
30	48	5.5	681	1	US-07-929-198-5	Sequence 5, Appl1
31	47.6	5.5	734	1	US-08-650-129-1	Sequence 1, Appl1
32	47.6	5.5	734	3	US-08-984-417-1	Sequence 1, Appl1
33	47.6	5.5	821	1	US-08-650-129-2	Sequence 2, Appl1
34	47.6	5.5	821	3	US-08-984-417-2	Sequence 2, Appl1
35	47.6	5.5	866	1	US-08-650-129-3	Sequence 3, Appl1
36	47.6	5.5	866	3	US-08-984-417-3	Sequence 3, Appl1
37	46.8	5.4	782	4	US-09-229-151C-9	Sequence 9, Appl1
38	46.8	5.4	833	4	US-09-229-151C-10	Sequence 10, Appl1
39	46.8	5.4	839	4	US-09-229-151C-8	Sequence 8, Appl1
40	46.8	5.4	884	2	US-08-851-974-2	Sequence 2, Appl1
41	46.8	5.4	884	2	US-09-213-390-2	Sequence 2, Appl1
42	46.8	5.4	2270	4	US-09-229-151C-13	Sequence 13, Appl1
43	46.8	5.4	2570	4	US-09-229-151C-14	Sequence 14, Appl1
44	46.4	5.4	693	4	US-09-402-515A-17	Sequence 17, Appl1
45	46.4	5.4	957	1	US-08-684-862-11	Sequence 11, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-620-312D-482  
; Sequence 482, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungang  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: Polypeptides  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 482  
; LENGTH: 866  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (42)..(749)  
; US-09-620-312D-482

Query Match 100.0%; Score 866; DB 4; Length 866;  
Best Local Similarity 100.0%; Pred. No. 7e-269;  
Matches 866; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTACTCCCTGAGCTAAGGGGAGAGAGCTGATCCATGAAATATGCTTCTATT 60  
DB 1 GGCATTACTCCCTGAGCTAAGGGGAGAGAGCTGATCCATGAAATATGCTTCTATT 60

QY	6	TGGGTGCTCGCTGGGAACTTTTCTTGCGACTCACTGTTCAAGAAAGAACCTG	120
Ds	61	TGGGTGCTCGCTGGGAACTTTTCTTGCGACTCACTGTTCAAGAAAGAACCTG	120
QY	121	CTCCATATTGGTGTACTCAAGTCTCATCTTCAACCCCTGTGTGGCGCTCTCATCAAC	180
Ds	121	CTCCATATTGGTGTACTCAAGTCTCATCTTCAACCCCTGTGTGGCGCTCTCATCAAC	180
QY	181	CCAGCTGGGTGCTGGGCCCACTCATCTGTATTATCCAAATCTGAAAGTATGCTGGAA	240
Ds	181	CCAGCTGGGTGCTGGGCCCACTCATCTGTATTATCCAAATCTGAAAGTATGCTGGAA	240
QY	241	ATTTCAGAGACAGAGTCAGAGACGGTACGTAAAGACAGACATTAACCCCATTCAGATCGTC	300
Ds	241	ATTTCAGAGACAGAGTCAGAGACGGTACGTAAAGACAGACATTAACCCCATTCAGATCGTC	300
QY	301	GCTACTGGAATACAGTCAATAGCGGCCCAAGAGATGACTCATGCTCATCAAGCTGGCTA	360
Ds	301	GCTACTGGAATACAGTCAATAGCGGCCCAAGAGATGACTCATGCTCATCAAGCTGGCTA	360
QY	361	AGCTGCGCATGCTCATGCCAAATGTCAGCCCTTCCCTGGCCACCAACCAATGTCAGGC	420
Ds	361	AGCTGCGCATGCTCATGCCAAATGTCAGCCCTTCCCTGGCCACCAACCAATGTCAGGC	420
QY	421	CAGGCACTGTCGTCTACTCTCAGGTTTGGACTGGAGCCAAGAAAAGTGGCCGACAC	480
Ds	421	CAGGCACTGTCGTCTACTCTCAGGTTTGGACTGGAGCCAAGAAAAGTGGCCGACAC	480
QY	481	CTGACTTGGCGGAGAACTGGAGGCCCGCGTATGTCGATCGAATGCCAAAAACAG	540
Ds	481	CTGACTTGGCGGAGAACTGGAGGCCCGCGTATGTCGATCGAATGCCAAAAACAG	540
QY	541	AACAGGAAAAAGCCACAGAAATTCCTATGATGTAATTTGTGAAAGTATTCAGCCGAA	600
Ds	541	AACAGGAAAAAGCCACAGAAATTCCTATGATGTAATTTGTGAAAGTATTCAGCCGAA	600
QY	601	TTTTTGGGAGGTGGCGGTGCTACTGTCTATCTGCAAGAACAGCTCCAGGAAATCGAGG	660
Ds	601	TTTTTGGGAGGTGGCGGTGCTACTGTCTATCTGCAAGAACAGCTCCAGGAAATCGAGG	660
QY	661	TGGGCACTTCATGGAGGGGACGTGGCATTAACCAATGTTTCAAAATATGTATCTT	720
Ds	661	TGGGCACTTCATGGAGGGGACGTGGCATTAACCAATGTTTCAAAATATGTATCTT	720
QY	721	GGAATTGAGAACTGCTTAAGGACAAGTGAACCTTCTTCCCTCTGCATTTCCACTGGC	780
Ds	721	GGAATTGAGAACTGCTTAAGGACAAGTGAACCTTCTTCCCTCTGCATTTCCACTGGC	780
QY	781	TCTGCATGACTATACAGACAGATTAATTTCCCTATTCGAAATAAATTCGAAATG	840
Ds	781	TCTGCATGACTATACAGACAGATTAATTTCCCTATTCGAAATAAATTCGAAATG	840
QY	841	AAAAATTGGGAATGTGCATCAATAGT	866
Ds	841	AAAAATTGGGAATGTGCATCAATAGT	866

```

RESULT 2
US-09-796-110-1
: Sequence 1, Application US/09796110
: Patent No. 6468776
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: TITLE OF INVENTION: Human Serine Protease
: FILE REFERENCE: 00-16
: CURRENT APPLICATION NUMBER: US/09/796,110
: CURRENT FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 3.0.
: SEQ ID NO 1
: LENGTH: 865
: TYPE: DNA
: ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)... (746)
US-09-796-110-1

```

Query Match	98.2%	Score 850.8;	DB 4;	Length 865;
Best Local Similarity	99.7%	Pred. No. 5.5e-264;		
Matches 863;	Conservative 0;	Mismatches 2;	Indels 1;	Gaps 1;

Qy	1	GGCACTTAATCTCCCTGACCTAAAGGGGGAAAGAGCTGGATCAACATGAAATATCTCTAT	60
Db	1	GGCACTTAATCTCCCTGACCTAAAGGGGGAAAGAGCTGGATCAACATGAAATATCTCTAT	60
Qy	61	TGGGTGTCTTCGCTGGGACATTTTTCCTTGTGACTCATCTGTTCAGAAAGAACCTGT	120
Db	61	TGGGTGTCTTCGCTGGGACATTTTTCCTTGTGACTCATCTGTTCAGAAAGAACCTGT	120
Qy	121	CTCCCTAATTTGGTGTACCTCAAGTCTCACTTCAACCCCTGTGTGGGGCGTCTCATCAAC	180
Db	121	CTCCCTAATTTGGTGTACTCTCAAGTCTCACTTCAACCCCTGTGTGGGGCGTCTCATCAAC	180
Qy	181	CCAGCTG3GTGTGGCCCCAGCTCACTGTCTATTACCAATCTGAAGTGAATGCTGGAA	240
Db	181	CCAGCTG3GTGTGGCCCCAGCTCACTGTCTATTACCAATCTGAAGTGAATGCTGGAA	240
Qy	241	ATTTCAAGACAGAGCTGAGAGCGGTACTGAACAGACAAATTAACCCCTTAGATCGGC	300
Db	241	ATTTCAAGAGAGAGTCAAGACG3GTACTGAACAGACAAATTAACCCCTTAGATCGTC	300
Qy	301	GCTACTGGAACTACAGTCAATAGCCGCCCAACAGATGACCTCATCAAGCTGGCTA	360
Db	301	GCTACTGGAACTACAGTCAATAGCCGCCCAACAGATGACCTCATCAAGCTGGCTA	360
Qy	361	AGCTGCCATGTCTAAATCCCAAGTCCAGCCCTTCCCTTGGCCACCAATGTCAAGC	420
Db	361	AGCTGCCATGTCTAAATCCCAAGTCCAGCCCTTCACTCGCCACCAACATGTCAAGC	420
Qy	421	CAGGCACTGTCTGTCTACTCTCAGGTTTGGACTGGAAGCAAGAAACAGTGGCGCACCC	480
Db	421	CAGGCACTGTCTGTCTACTCTCAGGTTTGGACTGGAAGCAAGAAACAGTGGCGCACCC	480
Qy	481	CTGACTTGC3GCAGAACCTGAGGCCCCCGTGATGTCTGATCGAATGCCAAAGAACAG	540
Db	481	CTGACTTGC3GCAGAACCTGAGGCCCCCGTGATGTCTGATCGAAGATGCCAAAGAACAG	540
Qy	541	AACAGAGAAAAGCCACAGAAATCTTATGTGTGAATTTGTGAAGTATTCAGCCGAA	600
Db	541	AACAGAGAAAAGCCACAGAAATCTTATGTGTGAATTTGTGAAGTATTCAGCCGAA	600
Qy	601	TTTTTGGGGAAGGTGGCGTGTCTACTGTCTCATCTGCAAGACCAAGCAAGCTCCAGGAATCGAGG	660
Db	601	TTTTTGGGGAAGGTGGCGTGTCTACTGTCTCATCTGCAAGACCAAGCAAGCTCCAGGAATCGAGG	660
Qy	661	TGGGGCACTTCATGTGGAGGGGAGCGTGGCATCTACACCAATGTTTACCAATATGTATCT	720
Db	661	TGGGGCACTTCATGTGGAGGGGAGCGTGGCATCTACACCAAGTGTTACCAAAATGTATCT	720
Qy	721	GGATTGGAACAACGTCTAAGGACAAAGTGAAGCCCTAATCTTCCCTCTGTCACTGGC	780
Db	721	GGATTGGAACAACGTCTAAGGACAAAGTGAAGCCCTAATCTTCCCTCTGTCACTTCCACT- GC	779
Qy	781	TCTGCAATGACATATACAGCAGATTAATTTTCCCTATTTCAAAATATAAATCTCCAAATG	840
Db	780	TCTGCAATGACATATACAGCAGATTAATTTTCCCTATTTCAAAATATAAATCTCCAAATG	839
Qy	841	AAAATTTGGGAAGTACATATCTGT 866	
Db	840	AAAATTTGGGAAGTACCAATCTAGT 865	

RESULT 3  
US-09-280-116-63/c  
; Sequence 63, Application US/09280116A

```
/ Patent No. 6331427
/ GENERAL INFORMATION:
/ APPLICANT: Robison, Keith E.
/ TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
/ FILE REFERENCE: 5800-24, 035800/176965
/ CURRENT APPLICATION NUMBER: US/09/280,116A
/ NUMBER OF SEQ ID NOS: 268
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 63
/ LENGTH: 865
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-63

Query Match          96.3%; Score 834; DB 4; Length 865;
Best Local Similarity 98.7%; Pred. No. 1,4e-258;
Matches 851; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 GGCACCTACCTCCCTGAGCTAAAGGGGAGAGAGCTGATCACCATTGAATATGCTCTATT 60
DB 865 GGCACCTACCTCCCTGAGCTAAAGGGGAGAGAGCTGATCACCATTGAATATGCTCTATT 806
QY 61 TGGGTGCTCCCTGCGGAGCAATTTTCTTGTGATCATCTGTTCAAGAAAGAGCCCTG 120
DB 805 TGGGTGCTCCCTGCGGAGCAATTTTCTTGTGATCATCTGTTCAAGAAAGAGCCCTG 746
QY 121 CTCCTATTTTGTGTGATCTCAAGTCTCACTCAACCCCTGTGTGGCGCTCATCAAA 180
DB 745 CTCCTATTTTGTGTGATCTCAAGTCTCACTCAACCCCTGTGTGGCGCTCATCAAA 686
QY 181 CCAGCTGGGTGCTGGCCCGCCAGCTCATCTGCTATTACCAATCTGAAAGTATCTGGAA 240
DB 685 CCAGCTGGGTGCTGGCCCGCCAGCTCATCTGCTATTACCAATCTGAAAGTATCTGGAA 626
QY 241 ATTTCAAGAGCAGAGTCAAGAGCGGTACTGAAACAGCAATTAACCCATTGAGATGCTC 300
DB 625 ATTTCAAGAGCAGAGTCAAGAGCGGTACTGAAACAGCAATTAACCCATTGAGATGCTC 566
QY 301 GCTACTGGAATCAAGTATAGCGCCCAAGAGATGACCTCATGCTCATCAAGCTGGCTA 360
DB 565 GCTACTGGAATCAAGTATAGCGCCCAAGAGATGACCTCATGCTCATCAAGCTGGCTA 506
QY 361 AGCTTCCGATGCTCAATCCCAAGTCCAGCCCTT-CCCTGCCACACCAATGTCAGG 419
DB 505 AGCTTCCGATGCTCAATCCCAAGTCCAGCCCTT-CCCTGCCACACCAATGTCAGG 446
QY 420 CCAGGCACTGTCTGTACTCTCAGGTTTGGATGGAGCCCAAGAAACAGTGGCCGACAC 479
DB 445 CCAGGCACTGTCTGTACTCTCAGGTTTGGATGGAGCCCAAGAAACAGTGGCCGACAC 386
QY 480 CCTGACTTGGGCGAGAACTGAGAGCCCGCTGATGTCTGATGAGAAATGCCAAAAACA 539
DB 385 CCTGACTTGGGCGAGAACTGAGAGCCCGCTGATGTCTGATGAGAAATGCCAAAAACA 326
QY 540 GAAACAAGAAAAAGCAACAGAAATTCCTTATGTGTGAATTTGTGAAGTATTCAGCCGA 599
DB 325 GAAACAAGAAAAAGCAACAGAAATTCCTTATGTGTGAATTTGTGAAGTATTCAGCCGA 266
QY 600 ATTTTGGGGAGGTTGGCCGCTTGTACTGTCACTGTGAAAGACAAAGCTCCAGGAAATCGAG 659
DB 265 ATTTTGGGGAGGTTGGCCGCTTGTACTGTCACTGTGAAAGACAAAGCTCCAGGAAATCGAG 206
QY 660 GTGGGGCACTTCAAGGAGGAGAGCTGCGCATCTACCAATGTTTAAATATATATATCC 719
DB 205 GTGGGGCACTTCAAGGAGGAGAGCTGCGCATCTACCAATGTTTAAATATATATATCC 146
QY 720 TGGATTGAGAACCTGCTTAAGAGCAAGTGAACCTTATCTTCCCTGCAATTCACCTGG 779
DB 145 TGGATTGAGAACCTGCTTAAGAGCAAGTGAACCTTATCTTCCCTGCAATTCACCTGG 86
```

```
QY 780 CTCCTGCCATGACTATACAGCAGATATATTTCTCTATTTCAAAAATATCTCCAAAT 839
DB 85 CTCCTGCCATGACTATACAGCAGATATATTTCTCTATTTCAAAAATATCTCCAAAT 26
QY 840 GAAATTTGGGAATGTAGCATA 861
DB 25 GAAATTTGGGAATGTAGCATA 4

RESULT 4
US-09-796-110-3
/ Sequence 3, Application US/09796110
/ Patent No. 6468776
/ GENERAL INFORMATION:
/ APPLICANT: Konklin, Darrell C.
/ TITLE OF INVENTION: Human Serine Protease
/ FILE REFERENCE: 00-16
/ CURRENT APPLICATION NUMBER: US/09/796,110
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 3
/ LENGTH: 705
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: This degenerate nucleotide sequence encodes the
/ OTHER INFORMATION: amino acid sequence of SEQ ID NO12.
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(705)
/ OTHER INFORMATION: n = A,T,C or G
US-09-796-110-3

Query Match          59.0%; Score 511.2; DB 4; Length 705;
Best Local Similarity 59.3%; Pred. No. 1.2e-154;
Matches 418; Conservative 163; Mismatches 124; Indels 0; Gaps 0;

QY 42 ATGAAATATGTCCTTATTTGGGTGCTCGCTGGGAGCAATTTTCTTGTGCTGACTATCT 101
DB 1 ATGAAATATGTCCTTATTTGGGTGCTCGCTGGGAGCAATTTTCTTGTGCTGACTATCT 60
QY 102 GTTCAGAAAGAGACCTCTCTCTATTTGTGTGTAAGTCAAGTCACTCAACCCCTGT 161
DB 61 GTTCAGAAAGAGACCTCTCTCTATTTGTGTGTAAGTCAAGTCACTCAACCCCTGT 120
QY 162 GTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCAAGCTCATGCTATTTACCAAT 221
DB 121 GTGGGCGTCTCATCAACCCAGCTGGGTGCTGGCCCAAGCTCATGCTATTTACCAAT 180
QY 222 CTGAAAGTATCTGGAATTTTCAAGACAGTCAAGAGCGTACTGAAACAGCAATTT 281
DB 181 CTGAAAGTATCTGGAATTTTCAAGACAGTCAAGAGCGTACTGAAACAGCAATTT 240
QY 282 AACCCCATTCAGATCGTCCGCTACTGAGAACTCAAGTCATAGGCCCCACAGATGACTC 341
DB 241 AACCCCATTCAGATCGTCCGCTACTGAGAACTCAAGTCATAGGCCCCACAGATGACTC 300
QY 342 ATGCTCATCAAGCTGCTTAAGCCCTGCTCAATGCCAAAGTCCAGGCCCTTCCCTC 401
DB 301 ATGCTCATCAAGCTGCTTAAGCCCTGCTCAATGCCAAAGTCCAGGCCCTTCCCTC 360
QY 402 GCCACCAACCAATGTCAGGCGCAGGCACTGCTCTACTCTCAGGTTTGGATCGAGCCAA 461
DB 361 GCCACCAACCAATGTCAGGCGCAGGCACTGCTCTACTCTCAGGTTTGGATCGAGCCAA 420
QY 462 GAAACAGTGGCGGACACCTGACTTGGCGGAGAACTGAGAGCCCGCTGATGTCTGAT 521
DB 421 GAAACAGTGGCGGAGACCTGACTTGGCGGAGAACTGAGAGCCCGCTGATGTCTGAT 480
QY 522 CGAAGATGCAAAAAACAGAACAGAAAAAGCCAACAGAAATCTCTTATGTGTGAATTT 581
DB 481 MNGARTGVCABABACAGARCARGNABRWSNCAYMSVNTTNGYGTNABRTTY 540
```

QY 582 GTGAAGATTAATGCGCAATTTTGTGGGAGGGGCGCGTGTCTGTGCATCTCCAGAGAC 6411

Db 541 GTMAARGATNTTWTSMNGNAHTHTTGGNGARGTNGCNGTNGCNA CNGTNAHTHTTAAAGAY 600

QY 642 AAGCTCCAGGAATCGAGGTGGGCGACTTCATGGAGGGGAGCGTCGCAATCTACCAAT 701

Db 601 AARYTNCARGAATTHGARGTNGNCAYTTTATGGGNGAGAYGTNGGNATHTYAACNAAY 660

QY 702 GTTTTCAATATGTATCCTGATTGAGAACACTGCTAAAGACAAG 746

Db 661 GTNTATYAARTAYGTWMSNTGATHGARAAYACNGCAAAAGATYAR 705

```

RESULT 5
US-09-120-582-1
: Sequence 1, Application US/09120582
: Patent No. 6087558
: GENERAL INFORMATION:
: APPLICANT: Howard, John A.
: APPLICANT: Hood, Elizabeth
: TITLE OF INVENTION: COMMERCIAL PRODUCTION OF PROTEASES IN PLANTS
: FILE REFERENCE: 10014
: CURRENT APPLICATION NUMBER: US/09/120,582
: CURRENT FILING DATE: 1998-07-22
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 825
: TYPE: DNA
: ORGANISM: Bos taurus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (44)..(730)
: US-09-120-582-1

```

Query Match	16.7%	Score 145:	DB 3:	Length 825:
Best Local Similarity	52.5%	Pred	No. 1.3e-36:	
Matches 347;	Conservative	0;	Mismatches 305;	Indels 9;
				Gaps 1;

Oy	111	GAAGACCAGCTCCCTAAATTTGGTATCTCAAGATCTCAATCAACCCCTGGTGGAGTCT	170
Db	86	GCAATATCTGTCCCTTACCAAGTGTCCCTGAATCTGGGTACCACTTCTGGGGGGCTCC	145
Oy	171	CTCATCAAAACCAAGCTGGGTGTGGGCCCAGCTCACTGTCTATTACAAATCTGAAAGTG	230
Db	146	CTCATCAACAGCCAGTGGGTGGTGTCTGGGGCTCATCTGCTACAAGTCCGGAATCCAAAGTG	205
Oy	231	ATGCTGGGAATTTCAAGACAGAGTCAAGACGGTACTGAAACAGCAATTAAACCCCATTT	290
Db	206	CGTCGGGAGAGAACAACTTATATGTCTGTAGGGGCAATGAGCAATTATCATAGCGGATCC	265
Oy	291	CAGATCGTCCGGCTACTGGAATACAGTCAATAGCGGCCCAAGGATACCTCATGCTCATC	350
Db	266	AAGATATCTGTCAATCCAGCTACAACTCAACACTTAAACAAGACATCATGCTGATT	325
Oy	351	AAAGTGGCTAAAGCTTGGCCAATGCTCAATCCAAAGTCAAGCCCTTCCCTCGCACCAACC	410
Db	326	AAACTGAAATCAGCTGCCAGTCTCAACAGCCAGTGGCCCTATCTCTGCGCAACATCC	385
Oy	411	AATGTCAGAGCCAGGACGTGTCTGTACTCTCAGGTTTGAATCGAGCCCAAGAAAACAT	470
Db	386	TGTGCTCTGCTGGGCAACCAAGTGTCTCATCTCTGGGTGGGGCAACACAAAGACATGGC	445
Oy	471	GAGCGACACCCCTGACTCTGGGGCAGAACTGGAAGGCCCCCGTGAATGTCTGATCGAATGC	530
Db	446	ACCACTTACCCCTGAAGTCTCTGAAGGTCTGAAAGGCTCCATCCTATCAGACAGCTCTTG	505
Oy	531	CAAAAAACAGAACAGGAAAAAGCCACAGAGAATCTCTTATGTGTGAATTTTGGAAAGTA	590
Db	506	AAAAGTGTCTACCCAGGCCAAGTCAACACAGCAACATTTCTGTGCGGGCTACCTCGAAGGGC	565
Oy	591	TTACGCCGAATTTTGGGGAGGTGGCCGTTGCTACTGTCACTGTGCAAAAGCAAGCTTCAG	650

Db	566	GGAAGAAGCTCCTGCCAGGTTGACTCCGATGGCCCTGTGTTCTGCAGTGGAAAGCTCCAG	625
Qy	651	GGATTCGAGTGGGGCACTTCATGGGAGGGGACGTGGCA-----TTTACACCAAT	701
Db	626	GGCATTTGTCTCCTGGGGGCTCTGGGCTGGCCCTCGAATAAACAGCCGTGTTCTTACACCAAG	685
Qy	702	GTTTACAAATATGTATCCTGGATTGAGAACACTGTGTAAGAGCAAGTGAGACCTACTTCT	761
Db	686	GTTCTCAACTACGTGAGCTGGATTAAAGACAGCACTACGCTCCCAACTAAATTAAGTTCATCT	745
Qy	762	C 762	
Db	746	C 746	

```

1      RESULT 6
2      US-08-956-267A-1
3      : Sequence 1, Application US/08956267A
4      : Patent No. 5945328
5      : GENERAL INFORMATION:
6      : APPLICANT: WOLDISEN, Heile Fabricius
7      : APPLICANT: KJELDSEN, Thomas Borglum
8      : TITLE OF INVENTION: A Process For Producing Trypsin
9      : TITLE OF INVENTION: (Trypsinogen)
10     : NUMBER OF SEQUENCES: 6
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: No. 59453280 No. 5945328disk of No. 5945328th America, Inc.
13     : STREET: 405 Lexington Avenue
14     : CITY: New York
15     : STATE: NY
16     : COUNTRY: USA
17     : ZIP: 10174
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Diskette
20     : COMPUTER: IBM Compatible
21     : OPERATING SYSTEM: DOS
22     : SOFTWARE: Fastseq for windows Version 2.0
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/956.267A
25     : FILING DATE: 22-OCT-1997
26     : CLASSIFICATION: 435
27     : ATTORNEY/AGENT INFORMATION:
28     : NAME: Rozek, Carol
29     : REGISTRATION NUMBER: 36,993
30     : REFERENCE/DOCKET NUMBER: 4500.204-US
31     : TELECOMMUNICATION INFORMATION:
32     : TELEPHONE: 212-867-0123
33     : TELEFAX: 212-878-9655
34     : INFORMATION FOR SEQ ID NO: 1:
35     : SEQUENCE CHARACTERISTICS:
36     : LENGTH: 897 base pairs
37     : TYPE: nucleic acid
38     : STRANDEDNESS: single
39     : TOPOLOGY: linear
40     : MOLECULE TYPE: cDNA
41     : FEATURE:
42     : NAME/KEY: CDS
43     : LOCATION: 4..744
44     : US-08-956-267A-1

```

	Query Match	16.1%	Score 139.6	DB 2	Length 897
	Best Local Similarity	52.0%	Pred. No. 7.5e-35		
	Matches 344	Conservative 0	Mismatches 309	Indels 9	Gaps 1
Qy	110	AGAAAGCCCTGCTCCCTATTGTGTGTACCTCAAGCTCATTTCAACCCCTGTGTGGGCGT	169		
Db	99	AGCAAAATTCATTCCCTACCAAGAGTGCTCCGTAAATTCTGGCTCCCACTTGTGTGGGCGT	158		
Qy	170	CCTCATCAAAACCAGCTGGGTGTGGGCCCAAGCTCACTGCTATTATTAACCAATCTGAAGT	229		
Db	159	CCTCATCAACAGCCAGTGGGTGTGTGTGCTGCTCACTGCTACAAGTCCGAATCCAGGT	218		
Qy	230	GATGCTGGGAAATTTCAAGACGACAGTCCAGACGCTACTGAACAGACATTAACCCCAT	289		

Db 219 GCCTCTGGGAGAACACAACTCCACGCTCTTGAGGGCAATGACCAATTCTCATATCCCGC 278  
Qy 290 TCAGATCTCCGCTACTGGAACATACAGTATAGCCGCCACAGATGACCTCTGCTCAT 349  
Db 279 CAAGATCATACACCACCCCAATTCATGAAATACCTTAAGAACATCATGCTGAT 338  
Qy 350 CAAGCTGGCTAAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTGCCACAC 409  
Db 339 TAACTGAGCTGACCTGCCACTCTCAACAGTCCAGTACATGCTCTCACTGCCAAGATC 398  
Qy 410 CAATGTCAGGCGCAGGACCTGCTGCTCTACTCTCAAGTTTGAGCTGAGCCAGAAAACAG 469  
Db 399 TTGTGAGCTGCTGTAACGAGTGTCTCATCTCTGGCTGGGGCAACACCAAGCAGTGG 458  
Qy 470 TGGCCCAACCCCTGACTTGGGCGAAGACCTGAGGCCCTGATGCTGATGAGAAATG 539  
Db 459 CTCACACTCCCTTCTGCTCTGCAATGCTGAAAGGCCCTGCTAAGTACAGTTCTTG 518  
Qy 530 CCAAAAACAGAACAGAAAGAAAGCCACAGAAATTCCTTATGTTGAAATTTGTAAGT 589  
Db 519 CAAGATCTCTACCCAGGCGCAGATCACCGGAAACATGATCTGTGTGGCTTCTGAGGG 578  
Qy 590 ATTACCGCAATTTTGGGGAGGTGGCCGTTGCTACTGTCATCTGCAAGACAGCTCCA 649  
Db 579 TGGTAAGATTTCTTGCAGGAGACCTGCTGGGCCCGCTGCTGCAATGACAGCTCCA 638  
Qy 650 GGGAAATCGAGTGGGCACTT-----CATGGAGGGGAGCTGGGCTTACACCA 700  
Db 639 GGGTATATGCTCTTGGGGCTATGCTGCGCCCGCCAGAAAACACCTGGGGCTTACACCA 698  
Qy 701 TGTTCATAATATGATCTCTGATTTGAAACACTGCTGAAAGAGAGACCTTCTTC 760  
Db 699 GGTCTCAACTATGTAATGATTCGATTCAGCAGACCAATCGCTGCCAATAAGATTCAAT 758  
Qy 761 TC 762  
Db 759 TC 760

RESULT 7  
US-09-205-258-189  
; Sequence 189, Application US/09205258  
; Patent No. 6525174  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/09/205,258  
; EARLIER FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 189  
; LENGTH: 1292  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-205-258-189

Query Match 7.6%; Score 66; DB 4; Length 1292;  
Best Local Similarity 47.1%; Pred. No. 4.9e-11;  
Matches 290; Conservative 1; Mismatches 301; Indels 24; Gaps 2;  
Qy 156 CCTGTGTTGGGCGCTCTCATCAACCAAGCTGGTGGTGGCCCAAGCTGCTGTAATTTA 215  
Db 342 CTCTGTGGGGCGACGCTATCGCCCCAGATGGCTCTGACAGCAGCCCACTGGCTCAAG 401



```

; NAME/KEY: SITE
; LOCATION: (36)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-247

```

```

Query Match
Best Local Similarity 46.9%; Pred. No. 6.1e-11;
Matches 289; Conservative 2; Mismatches 301; Indels 24; Gaps 2;

```

```

156 CCCTGTGTGGGCGCTCTCATCAAAACCCAGCTGGGTGTGGCCCACTCACTCTATTTA 215
196 CTCTGTGGGGCCAGCGTCAATCGCCCAAGTGGCTCTGTACAGACAGGCCCACTGCCAAG 255
216 CCAATCTGAAAGTATGCTGGGAAATTTCAAGAGCAGAGTCAAGACGGTACTGAACAG 275
256 CCCCCTAATAGTTTCACTGGGGGAGCAACCTCCAGAGAGGAGGGGCTGTAGACAG 315
276 ACAATTAAACCCCTTTCAGATCCGCTACTGGAATCAAGTATAGCCGCCCAAGAT 335
316 ACCGGACAGCCACTGATGCTTCCCAACCCGCTTCAACCAAGCCTCCCAACAA 375
336 GACC-----TCATGCTCATCAAGCTGAGCTGAGCTGCAATGCTCAATCCCAAA 383
376 GACCAACCGCAATGATCATGCTGTGAGATGAGTGCATGCCAGTCTTCATCACTGGGCT 435
384 GTCCAGACCCCTTCCCTCGCCACCAACATGTAGGCGCAGCACTGTGTCTACTCTCA 443
436 GTGCGACCCCTCACTCTCTCTCAGCTGTGTGATGCTGAGGACAGCTGTCTATTTC 495
444 GATTGTGACTGGAGCCCAAGAAACAGTGGCCGACACCTGTACTTGGGCGAAGACTGGAG 503
496 GGCTGGGGGAGAGAGTCCAGCCCCCAGTTTACGCTGCTCAACCTTGGATGCGCAAC 555
504 GCCCCCGTGTGCTGTATGAGATGCAAGATCCCAAAACAGAAAGAAAGCCACAGAAAT 563
556 ATACCATCATTTAGACACCAAGAGTGAAGACGCTTACCCCGGCAATCAAGACACC 615
564 TCCTTATGTGTGAATTTGTGAAGATTCAGCCGAATTTTGGGAGAGTGGCGCTTGTCT 623
616 ATGGTGTGTGCGCAGCGTGAAGAGAGGGGGCAAGAGACTCCGCAAGGTATCTCGGGGGG 675
624 ACTGTATCTGCAAAAGCAAGCTCCAGGGAAAT-----CGAGTGGGGCACTTC 671
676 CCTGTGTGTGTAAACAGTCTCTTCAAGGCATTATCTCTGGGGCCAGATCCGTGTGCG 735
672 ATGGGAGGGAGCGTGGCATCTTACCAATGTTTAAAGAAATGATTCCTGGATTGAAGAC 731
736 ATACCCGAAAGCGTGTGTCTTACGAAAGTCTGCAAAATGTGAATGATTCAGAGAG 795
732 ACTGCTAAGGACAAGT 747
796 AGCATGAAGACAATT 811

```

```

RESULT 9
US-09-280-116-64
; Sequence 64, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 452
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-64

```

```

Query Match
Best Local Similarity 52.0%; Pred. No. 5.4e-11;
Matches 196; Conservative 0; Mismatches 175; Indels 6; Gaps 2;

```

```

15 GAGCTAAGGGGAGAGAGCTGATCAACCAAGAAATAGTCTTATTTGGGTGCTCCGT 74
75 GAGCTAAGCTGGGCTTTCAACATCATCAATGAAGTTATCTCTGCGGCTCTGAAT 134
75 GGAACATTTTCTTCTGCTGACTCATCTGTTCAGAAAGAACCCCTCCCT--TATTGG 131
135 CTGACGTGTGCTTTGGCTTTTATTCAGATTTACAGTCAAGCTCCACCTCCCTTACTTG 194
132 GTGTACTCAAGTCTCACTTCAACCCCTGTGGGGGCTCTCATCAACCCAGCTGGGTG 191
195 GTCTATTTGAATCTGACTACTTGGCCCTGCGTGGAGTCTGTATCCACCGCTTGGGTG 254
192 CTGGCCCAAGCTCATCTGCTATTTCAGAAATCTGAAAGTATGCTGGGAAATTTCAAGAGC 251
255 ATCAACAGCTGCACACTGCAATTTACCAAGCTTGGGTATATTGGGGGTTCATCAATCCA 314
252 AGAGTGAAGAGGGTACTGACAGACAA--ATTAAACCCATTAGATGCTCGCTACTGG 308
315 GCAAGCTTAAATGAAGAGATTTGCAAGTATGGCTTATGAGAAAGTATTCATTCCTCA 374
309 AACTACAGTATAGGCGCCCAAGATGATGATCTATGCTCATCAAGCTGCTAAGCTGCGC 368
375 CACTTCTGACGACTCTTATTTGATATGATGATGATGATGATGATGATGATGATGATGAT 434
369 ATGCTCAATCCCAAGT 385
435 GAATCAATGACTATGT 451

```

```

RESULT 10
US-09-386-642-10
; Sequence 10, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-10

```

```

Query Match
Best Local Similarity 47.1%; Pred. No. 1e-10;
Matches 290; Conservative 0; Mismatches 302; Indels 24; Gaps 2;

```

```

156 CCCTGTGTGGGCGCTCTCATCAAAACCCAGCTGGGTGTGGCCCACTCACTCTATTTA 215
241 CTCTGTGGGGCCAGCGTCAATCGCCCAAGTGGCTCTGTACAGACAGGCCCACTGCCAAG 300
216 CCAATCTGAAAGTATGCTGGGAAATTTCAAGAGCAGAGTCAAGACGGTACTGAACAG 275
301 CCCCCTAATAGTTTCACTGGGGGAGCAACCTCCAGAGAGGAGGGGCTGTAGACAG 360

```



QY 276 ACAATTACCCATTCAGATGTCCTGCTACTGCACTACAGTCAAGAGCCCAAGAT 335  
DB 361 ACCGCGACAGCACTGATGCTCTTCCACCCCGCTTCAACAACAGCTCCCAACAA 420  
QY 336 GACC-----TCATGCTCATGAGTGGCTAAGCTGCGCATGCTCAATCCCAA 383  
DB 421 GACCAACCGCATGATGATGCTGTAAGATGAGTGGCATGCGAGTCTTCATCATCCTGGGCT 480  
QY 384 GTCAGCCCTTCCCTTCCGCAACCAATGTCAGGCGAGGAGTGTCTGCTACTCTCA 443  
DB 481 GTGGAGACCCCTCACTCTCTCTCAAGCTGTGCTACTGCGACACAGCTGCTCATTTCC 540  
QY 444 GGTTCGATCTGAGCCCAAGAAACAGTGGCCGACACCTGACTTGGCGAGAACCTGAG 503  
DB 541 GGTGGGGGAGCAGTCCAGCCCGCATTAAGCTGCTGCTCACTTGCATGCGAGCCAC 600  
QY 504 GCGCCGCTGATGCTGATGCAAGATGCAAAACAAAGAAAGAAAGCAAGAT 563  
DB 601 ATCACCATCATGAGCAACGAGAGTGAAGACGCTTACCCCGCAATCATCAGACACC 660  
QY 564 TCCTTATGTGTAATTTGTGAAGTATTCAGCCGAATTTTGGGAGGTGGCCGTTGCT 623  
DB 661 ATGCTGTGTCAGCGTGCAGAGAGGGGCGAAGACTCTGCGAGGCTACTCCGGGGCC 720  
QY 624 ACTGTATCTGCAAGCAAGCTCCAGGGAAAT-----CGAGTGGGGCACTTC 671  
DB 721 CTTCTGCTGTGAACAGTCTCTTCAAGCATTAATCTCTGAGGCGACAGATCCGTGCG 780  
QY 672 ATGGAGAGGAGACGTCGGCATCTACACCAATGTTTAAATATGATCTGATGTAAC 731  
DB 781 ATCACCGAAGAGCTGTGTCTACAGAAAGTGTGCAAAATATGTGACTGATCGAGAG 840  
QY 732 ACTGCTAAGAGCAAGT 747  
DB 841 ACGATGAAGAACAAATT 856

RESULT 11  
US-08-944-483-7  
; Sequence 7, Application US/08944483  
; Patent No. 6232456

## GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE  
APPLICANT: COLETTIS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: KLAS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUPE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183, US. 01  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1166 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-944-483-7

Query Match 7.5%; Score 64.8; DB 3; Length 1166;  
Best Local Similarity 47.1%; Pred.No. 1.1e-10;  
Matches 290; Conservative 0; Mismatches 302; Indels 24; Gaps 2;

QY 156 CCTGTGTGGGCGTCTCATCAAAACCAAGCTGGGTGCTGGCCCACTCACTGCTATTTA 215  
DB 238 CTCTGTGGGCGAGCGCTCATTCGCCCAATGCTCTGTACAGCAACCACTGCTCAG 227  
QY 216 CCAATCTGAAGTATGCTGGGAAATTTCAAGAGCAGTCAAGACGGTACTGAACAG 275  
DB 298 CCGGCTACTAGTTCACCTGGGGAGCAACCTCAGAGAGAGAGGGCTGTGAGCAG 357  
QY 276 ACAATTACCCATTCAGATGTCCTGCTAGTGAATGATAGTATGATGCGCCCAAGAT 335  
DB 358 ACCGAGACGCCAAGTATGCTTCCCAACCCGCTTCAACAAGCTCTCCCAACAA 417  
QY 336 GACC-----TCATGCTCATCAAGCTGCTAAGCTGCGCATGCTATCCCAA 383  
DB 418 GACCAACCGCATGATGATGCTGTGAGATGATGATGATGATGATGATGATGATGATG 477  
QY 384 GTCAGCCCTTCCCTCCGCAACCAATGTCAGGCGAGCACTGTCTGTACTCTCA 443  
DB 478 GTGCGACCCCTCACTCTCTCTCAGCTGTGTCACTGTGGCAACAGCTGCTCATTTCC 537  
QY 444 GGTTCGATCTGAGCCCAAGAAACAGTGGCCGACACCTGATCTTGGGGAGAACCTGAG 503  
DB 538 GGTGGGGAGAGCGTCCAGCCCGCATTAAGCTGCTCTCAACCTTGGGATGGCCAC 597  
QY 504 GCGCCGCTGATGCTCATGAGATGCAAAACCAAGAAAGAAAGCAAGAT 563  
DB 598 ATCACCATCATGAGCAACGAGAGTGAAGAGCTTACCCGCAACATCAGACACC 657  
QY 564 TCCTTATGTGTAATTTGTGAAGTATTCAGCGAAATTTTGGGAGGTGGCGTTGCT 623  
DB 658 ATGCTGTGTCAGCGTGCAGAGAGGGGCAAGACTCTGCGAGGCTACTCCGGGGCC 717  
QY 624 ACTGTATCTGCAAGCAAGCTCCAGGAAAT-----CGAGTGGGGCACTTC 671  
DB 718 CTTCTGCTGTGAACAGTCTCTTCAAGCATTAATCTCTGGGGCCAGATCCGTGCG 777  
QY 672 ATGGAGGGAGCGTCCGATCTACACCAATGTTTAAATATGATCTGATGTAAC 731  
DB 778 ATCACCGAAGCGTGTGTCTTACAGAAAGTGTGAAATATGTGATGATTCAGAG 837  
QY 732 ACTGCTAAGAGCAAGT 747  
DB 838 ACGATGAAGAACAAATT 853

RESULT 12  
US-08-944-483-8  
; Sequence 8, Application US/08944483  
; Patent No. 6232456  
GENERAL INFORMATION:  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLETTIS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.

APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: KLAS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUBE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1192 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-944-483-8

Query Match 7.5%; Score 64.8; DB 3; Length 1192;  
Best Local Similarity 47.1%; Pred. No. 1.1e-10;  
Matches 290; Conservative 0; Mismatches 302; Indels 24; Gaps 2;

156 CCTGTGTGGGCTCTCATCAAAACCAAGTGGTGTGCGCCCACTCACTGCTATTTA 215  
242 CTCTGTGGGCGCAGCTCATCGCCCAAGTGGCTCTGACACACCCCACTGCTCAAG 301  
216 CCAATCTGAAGTGTGCTGGAATTTCAAGAGCAGAGTGAAGCGTACTGAACAG 275  
302 CCCCCTAATGATGTTACCTGGGGCAGACACCTCCAGAAAGAGAGGGCTGTAGCAG 361  
276 ACAATTAACCCATTGATGATGTCGCTACTGAACTACATGATAGCGCCCAAGAT 335  
362 ACCCGAGACGCACTAGTCTTCCCCCAACCGGCTTCAACAACAGCTCCCAACAA 421  
336 GACC-----TCATGTCATCAAGCTGGCTAAGCTGCCATGCTCATTCCTCAAA 383  
422 GACCAACCGAATGATCATGCTGTGAGATGAGTGCATGCGCACTTCATCACTGGGCT 481  
384 GTCCAGCCCTTCCCTCGCCACACCAATGTGAGGCGAGGCACTGTCTGCTACTCTCA 443  
482 GTGCGACCCCTTCTCTCTCAGCGTGTGTCATGCTGCGACGAGTCCCTCATTTTC 541  
444 GGTGTGACTGAGCCAGAAAGAAACAGTGGCCGACACCTGACTTGGGGGAGAACTGGAG 503  
542 GGCTGGGGGAGCAGTTCAGAGCCCGCAATTACGCTGCTCAACACTTGGGATGCGCAAC 601

504 GCCCCGATGATCTGATGAGAAATGCCAAAAACAGAAACAGAAACCAAGCAAGAT 563  
602 ATCACCATCATGATGAGACAGAAAGTGTGAGAACGCTTACCCGGCAACATCAGACACC 661  
564 TCCTATGTGTGAATTTGTGAAGTATTTCAGCCCAATTTTGGGAGGTGGCCGTGCT 623  
662 ATGTGTGTGCCAGCCTGAGAGAGGGGCAAGAGACTCTGCGCAGGTGATCTCCGGGGC 721  
624 ACTGATCTGCAAGACAGAGCTCCAGGAT-----CGAGTGGGGCACTTC 671  
722 CCTGTGTGTGAACAGTCTCTTCAAGCAATATCTCTGGGGCCAGATTCCTGTGCG 781  
672 ATGGAGGGGAGGTCGGCATCTTACACCAATGTTACAAATATGTATCTGATTTGAGAAC 731  
782 ATCACCAGAAAGCTGTGTCTTACACGAAGTCTGCAAAATGTGATGATCCAGAG 841  
732 ACTGTAAGACAGT 747  
842 ACGATGAAGACAAAT 857

RESULT 13  
US-09-025-059-2  
Sequence 2, Application US/09025059  
Patent No. 6075136

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom  
APPLICANT: Corley, Neil C.  
APPLICANT: Guejler, Karl J.  
TITLE OF INVENTION: PROSTATE-ASSOCIATED SERINE PROTEASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/025,059  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0481 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1314 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LUNGUT10  
CLONE: 2723646  
US-09-025-059-2

Query Match 7.5%; Score 64.8; DB 3; Length 1314;  
Best Local Similarity 47.1%; Pred. No. 1.2e-10;  
Matches 290; Conservative 0; Mismatches 302; Indels 24; Gaps 2;

```

QY 156 CCCTGTGTGGCGCTCTCATCAAAACCAGCTGGTGTGCGCCCGCTACCTGCTATTTA 215
D 359 CTCTGTGTGGCGCGAGCTCATCTGCCCCAGATGGCTCTGACAGAGCCACCTGGCTCAAG 418
QY 216 CCAATCTGAAATGATGCTGGGAAATTTTCAAGAGCAGTCAAGACGGTACTGAAACAG 275
D 419 CCCCCTACATAGTTCACCTGGGGCAGACAACTCCAGAGAGGAGGGCTGTGAGACAG 478
QY 276 ACAATTAACCCCAATTCAGATCTGTCGGCTACAGATCTACAGTCAAGCGCCCAAGAT 335
D 479 ACCCGAGACGCACTGATGCTCTTCCCGACCCCGCTTCAACAAGACCTCCCAACAA 538
QY 336 GACC-----TCATGCTCATCAAGCTGGCTGAAGCTGGCATGCTCAATCCCAAA 383
D 539 GACCACGCAATGACATCATGTGTGTAAGATGGCATGCGCAATCTCATCACTCGGGCT 598
QY 384 GTCCAGGCTCTTCCCTTCCGACCAACCAATGTCAAGGCGCAGCACTGTCTTACTCTCA 443
D 599 GTGCGACCTCTCACTCTCTCTCAAGCTGTGTCATCTGCGCACAGCTGCTCATTTCC 658
QY 444 GATTGGACTGAGCCAAAGAAACAGTGGCGAGACCTGACCTTGGCGGAGAACTGGAG 503
D 659 GCGTGGGGGAGCAGCTCACGCCCCAGTTAGCGCTGCTTCAACCTTGCATGCGCAAC 718
QY 504 GCCCGCGTATGTCATGATGAGATGCCAAAGAAACAGAGAGAAAGCCACAGGAAT 563
D 719 ATACCATCATTTAGAGACCAAGATGTGAGAGAGCTTACCAGGCAACATACAGACACC 778
QY 564 TCCTTATGTGTGAATTTTGTGAAGATTTTCAAGCCGAATTTTGGGGAGTGGCCGTGCT 623
D 779 ATGATGTGTCCAGCGTGCAGAGAGAGGGGAGAGAGCTCTGCGAGGCTGATCGGGGGC 838
QY 624 ACTGATCATGCAAGAGCAAGCTCCAGGGAAT-----CGAGTGGGGCACTTC 671
D 839 CTTCTGTCTGTAACAGCTCTTCAAGGCAATATCTCTGGGGGCAAGATCCGTGTGCG 898
QY 672 ATGGAGGGGAGCGTGGCATCTACCAATGTTTACAAATATGATCTGATTTAGAAC 731
D 899 ATACCCGAAAGCGTGGTGTCTACAGAAAGTCTGAAATATGTGACTGTGATCCAGGAG 958
QY 732 ACTGCTAAGAGCAAGT 747
D 959 ACGATGAGAAACAATT 974

```

```

RESULT 14
US-08-790-137-2
; Sequence 2, Application US/08790137
; Patent No. 5840871
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-ASSOCIATED
; TITLE OF INVENTION: KALIKREIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,137
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0195 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-4166
; TELEFAX: 415-845-0555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-790-137-2

```

```

Query Match 7.3%; Score 63.4; DB 2; Length 833;
Best Local Similarity 46.8%; Pred. No. 2, 6e-10;
Matches 288; Conservative 0; Mismatches 304; Indels 24; Gaps 2;

```

```

QY 156 CCCTGTGTGGCGCTCTCATCAAAACCAGCTGGTGTGCGCCCGCTACCTGCTATTTA 215
D 171 CTCTGTGTGGCGCGAGCTCATCTGCCCCAGATGGCTCTGACAGAGCCACCTGCTCAAG 230
QY 216 CCAATCTGAAATGATGCTGGGAAATTTTCAAGAGCAGTCAAGACGGTACTGAAACAG 275
D 231 CCCCCTACATAGTTCACCTGGGGCAGACAACTCCAGAGAGGAGGGCTGTGAGACAG 290
QY 276 ACAATTAACCCCAATTCAGATCTGTCGGCTACAGATCTACAGTCAAGCGCCCAAGAT 335
D 291 ACCCGAGACGCACTGATGCTCTTCCCGACCCCGCTTCAACAAGACCTCCCAACAA 350
QY 336 GACC-----TCATGCTCATCAAGCTGGCTGAAGCTGGCATGCTCAATCCCAAA 383
D 351 GACCACGCAATGACATCATGTGTGTAAGATGGCATGCGCACTTCCATCACTGGGCT 410
QY 384 GTCCAGGCTCTTCCCTTCCGACCAACCAATGTCAAGGCGCAGCACTGTCTTACTCTCA 443
D 411 GTGCGACCTCTCACTCTCTCTCAAGCTGTGTCATCTGCGCACAGCTGCTCATTTCC 470
QY 444 GATTGGACTGAGCCAAAGAAACAGTGGCGAGACCTGATCTTGGGGAGAACTGGAG 503
D 471 GCGTGGGGGAGCAGCTCACGCCCCAGTTAGCGCTGCTTCAACCTTGGCATGCGCAAC 530
QY 504 GCCCGCGTATGTCATGATGAGATGCCAAAGAAACAGAGAGAAAGCCACAGGAAT 563
D 531 ATACCATCATTTAGAGACCAAGATGTGAGAGAGCTTACCAGGCAACATACAGACACC 590
QY 564 TCCTTATGTGTGAATTTTGTGAAGATTTTCAAGCCGAATTTTGGGGAGTGGCCGTGCT 623
D 591 ATGATGTGTGCCAGCGTGCAGAGAGAGGGGCAAGAGACTCTGCCAGGGTGACTCCGGGGCG 650
QY 624 ACTGATCTGCAAGAGCAAGCTCCAGGGAAT-----CGAGTGGGGCACTTC 671
D 651 CTTCTGTCTGTAACAGCTCTTCAAGGCAATATCTCTGGGGGCAAGATCCGTGTGCG 710
QY 672 ATGGAGGGGAGCGTGGCATCTACCAATGTTTACAAATATGATCTGATTTAGAAC 731
D 711 ATACCCGAAAGCGTGGTGTCTACAGAAAGTCTGAAATATGTGACTGTGATCCAGGAG 770
QY 732 ACTGCTAAGAGCAAGT 747
D 771 ACGATGAGAAACAATT 786

```

```

RESULT 15
US-08-361-395-2
; Sequence 2, Application US/08361395
; Patent No. 5733768
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.

```

APPLICANT: No. 57337681s, Franklin H.  
 TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Eli Lilly and Company  
 STREET: Lilly Corporate Center/Patent Division  
 CITY: Indianapolis  
 STATE: Indiana  
 COUNTRY: United States  
 ZIP: 46285  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/361,395  
 FILING DATE: 22-DEC-1994  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Blalock, Donna K.  
 REGISTRATION NUMBER: 38,082  
 REFERENCE/DOCKET NUMBER: X8350A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 317-277-1090  
 TELEFAX: 317-276-3861  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 732 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-361-395-2

Query Match 7.3%; Score 63; DB 1; Length 732;  
 Best Local Similarity 53.4%; Pred. No. 3.2e-10;  
 Matches 132; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY	156	CCCTGTGTGGGCTCTCTCATCAACCAGCTGGGTGCTGGCCCAAGTCACTGCTATTTA	215
DB	136	CTCTGTGTGGGGGTCTTATCATCCATGCTGGGTCTCTCACAGCTGCCCAAAAAA	195
QY	216	CCAAATCTGAAAGTGAATGCTGGGAAATTTCAGAGCAGAGTCAGAGCGGTAAGTGAACG	275
DB	196	CCGAATCTTCAGGCTCTTCTGCGGGAAGCATTAACCTTCGCAAGGAGAGTTCCAGAG	255
QY	276	ACAAATTAACCCCATTCAGATCGTCCGCTACTGGAATCAAGTATAGCGCCCAAGAGAT	335
DB	256	CAGAGTTCGTGTGTCGGGGCTGTGATCCACCTGACTATGATGCCGCAAGCATGACAG	315
QY	336	GACCTCATGCTCATCAAGCTGCTAAAGCTGCAATGCTCAATCCAAAGTCCAGCCCTT	395
DB	316	GACATCATGCTGTGGGCTGGGACGCGCCAGCCAACTCTGAACTCATTCAGCCCTT	375
QY	396	CCCTCG 402	
DB	376	CCCTGG 382	

Search completed: November 7, 2003, 23:05:40  
 Job time : 84 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 23:04:21 ; Search time 351 Seconds  
(without alignments)  
7672.847 Million cell updates/sec

Title: US-10-037-270-482

Perfect score: 866  
Sequence: 1 ggcactactccctgagcta.....tgggaatgacatagctagt 866

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 214354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	866	100.0	866	14	US-10-037-270-482
2	862.8	99.6	1060	10	US-09-910-151-1
3	850.8	98.2	865	12	US-10-246-006-1
4	730	84.3	730	11	US-09-842-758-11
5	728.4	84.1	730	11	US-09-842-758-13
6	706.4	81.6	708	10	US-09-910-151-3
7	703.8	81.3	721	11	US-09-842-758-106
8	606.4	70.0	798	9	US-09-888-615-43
9	511.2	59.0	705	12	US-10-246-006-3
10	154.4	17.8	853	10	US-10-101-510-199
11	154	17.8	850	12	US-09-954-456-748
12	153.2	17.7	802	9	US-09-923-779-148
13	153.2	17.7	802	12	US-10-101-510-210
14	142	16.4	699	12	US-10-181-808-1
15	132.8	15.3	522	9	US-09-923-779-91
16	132.8	15.3	534	9	US-09-923-779-97

C 17	132.8	15.3	566	9	US-09-923-779-98	Sequence 98, App1
C 18	132.8	15.3	567	9	US-09-923-779-142	Sequence 142, App
C 19	132.8	15.3	572	9	US-09-923-779-21	Sequence 21, App1
C 20	132.8	15.3	592	13	US-10-040-739-1120	Sequence 1120, App
C 21	131.8	15.2	566	9	US-09-923-779-99	Sequence 99, App1
C 22	131	15.1	556	9	US-09-923-779-86	Sequence 86, App1
C 23	130.6	15.1	565	9	US-09-923-779-92	Sequence 92, App1
C 24	130.4	15.1	566	9	US-09-923-779-100	Sequence 100, App1
C 25	130.2	15.0	560	9	US-09-923-779-83	Sequence 83, App1
C 26	120.8	13.9	567	14	US-10-060-036-4206	Sequence 4206, App
C 27	102	11.8	957	12	US-10-170-789-51	Sequence 51, App1
C 28	102	11.8	957	13	US-10-045-367A-1	Sequence 1, App1
C 29	102	11.8	957	13	US-10-137-878-247	Sequence 247, App
C 30	102	11.8	1123	12	US-10-140-021-247	Sequence 247, App
C 31	102	11.8	1123	12	US-10-140-021-247	Sequence 247, App
C 32	102	11.8	1123	12	US-10-140-274-247	Sequence 247, App
C 33	102	11.8	1123	12	US-10-140-471-247	Sequence 247, App
C 34	102	11.8	1123	12	US-10-140-807-247	Sequence 247, App
C 35	102	11.8	1123	12	US-10-140-922-247	Sequence 247, App
C 36	102	11.8	1123	12	US-10-140-924-247	Sequence 247, App
C 37	102	11.8	1123	12	US-10-140-926-247	Sequence 247, App
C 38	102	11.8	1123	12	US-10-141-698-247	Sequence 247, App
C 39	102	11.8	1123	12	US-10-141-702-247	Sequence 247, App
C 40	102	11.8	1123	12	US-10-141-704-247	Sequence 247, App
C 41	102	11.8	1123	12	US-10-142-421-247	Sequence 247, App
C 42	102	11.8	1123	12	US-10-142-433-247	Sequence 247, App
C 43	102	11.8	1123	12	US-10-142-767-247	Sequence 247, App
C 44	102	11.8	1123	12	US-10-143-033-247	Sequence 247, App
C 45	102	11.8	1123	12	US-10-144-994-247	Sequence 247, App

#### ALIGNMENTS

RESULT 1  
US-10-037-270-482  
Sequence 482, Application US/10037270  
GENERAL INFORMATION: US20030104529A1

APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyang  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aifeng J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yungting  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillingshast, John  
APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. US20030104529A1 Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP28  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pc\_fl\_genes Version 1.0  
SEQ ID NO 482  
LENGTH: 866  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS

LOCATION: (42) . (749)  
US-10-037-270-482

Query Match 100.0%; Score 866; DB 14; Length 866;  
Best Local Similarity 100.0%; Pred. No. 2,76-278;  
Matches 866; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 GGCACCTACTCCCGAGCTAAGGAGGAGAGCTGATCACCATGAATATGCTTCTATT 60
Db 1 GGCACCTACTCCCGAGCTAAGGAGGAGAGCTGATCACCATGAATATGCTTCTATT 60
Qy 61 TGGGTGCTCTGCTGGAGCATTTTCTTTGCTGATCTGATCTGTTCAAGAGAGACCTTG 120
Db 61 TGGGTGCTCTGCTGGAGCATTTTCTTTGCTGATCTGATCTGTTCAAGAGAGACCTTG 120
Qy 121 CTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGTGTGGGGTCTCTATCAAC 180
Db 121 CTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGTGTGGGGTCTCTATCAAC 180
Qy 181 CCAGCTGGGTGTGGGCCCCAGCTCACTGCTATTATTAACCAATCTGAAGTGTGCTGGGAA 240
Db 181 CCAGCTGGGTGTGGGCCCCAGCTCACTGCTATTATTAACCAATCTGAAGTGTGCTGGGAA 240
Qy 241 ATTTCAAGAGAGAGTGAAGAGCGGTACTGAACAGACATTTAACCCATTCAGATCGTCC 300
Db 241 ATTTCAAGAGAGAGTGAAGAGCGGTACTGAACAGACATTTAACCCATTCAGATCGTCC 300
Qy 301 GCTACTGTGAAGTCAAGTCAATAGCCCCCAAGAGTACCTCACTGCTATCAAGTGGCTA 360
Db 301 GCTACTGTGAAGTCAAGTCAATAGCCCCCAAGAGTACCTCACTGCTATCAAGTGGCTA 360
Qy 361 AGCTGCTGATGCTCAATCCCAAGTCCAGCCCTTCCCTGAGCAGCAGCAATGTCAGGC 420
Db 361 AGCTGCTGATGCTCAATCCCAAGTCCAGCCCTTCCCTGAGCAGCAGCAATGTCAGGC 420
Qy 421 CAGGACGTGCTGTCTACTCTCAGGTTTGAAGTGAAGCAAGAAAGAGTGGCCGAGACC 480
Db 421 CAGGACGTGCTGTCTACTCTCAGGTTTGAAGTGAAGCAAGAAAGAGTGGCCGAGACC 480
Qy 481 CTGACTTGGCGGAGAACTTGAAGGCCCCCGGATGTCTGATCGAAGATGCCAAAAAAG 540
Db 481 CTGACTTGGCGGAGAACTTGAAGGCCCCCGGATGTCTGATCGAAGATGCCAAAAAAG 540
Qy 541 AACAAAGAAAAAGCCACAGAAATTCCTTATGTGAATTTGTGAAGTATTCAGCCGAA 600
Db 541 AACAAAGAAAAAGCCACAGAAATTCCTTATGTGAATTTGTGAAGTATTCAGCCGAA 600
Qy 601 TTTTGGGAGAGTGGCGGCTTGTCTACTGTCACTGCAAAAGACAGCTCAGGGAATCGAG 660
Db 601 TTTTGGGAGAGTGGCGGCTTGTCTACTGTCACTGCAAAAGACAGCTCAGGGAATCGAG 660
Qy 661 TGGGGCACTTCAATGAGGAGGAGCGTGGCATCTACCAATGTTTAAATATGTATCT 720
Db 661 TGGGGCACTTCAATGAGGAGGAGCGTGGCATCTACCAATGTTTAAATATGTATCT 720
Qy 721 GGATTGGAACACTGCTAAGGAGCAAGTGAAGCCTTCTCTCCCTGCAATTCACGAGC 780
Db 721 GGATTGGAACACTGCTAAGGAGCAAGTGAAGCCTTCTCTCCCTGCAATTCACGAGC 780
Qy 781 TCTGCACTGAGCTATACAGAGAGTAATTTTCCCTATTCAAATAATATCTCAATG 840
Db 781 TCTGCACTGAGCTATACAGAGAGTAATTTTCCCTATTCAAATAATATCTCAATG 840
Qy 841 AAAATTTGGGAATGTACATCTAGT 866
Db 841 AAAATTTGGGAATGTACATCTAGT 866
```

RESULT 2  
US-09-910-151-1  
Sequence 1, Application US/09910151  
Patent No. US20020137181A1  
GENERAL INFORMATION:

```
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 14087, A NOVEL SERINE PROTEASE MOLECULE
FILE REFERENCE: 38155-20021.00
CURRENT APPLICATION NUMBER: US/09/910,151
PRIOR FILING DATE: 2001-07-18
PRIORITY APPLICATION NUMBER: US 60/219,002
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1060
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS (164)...(871)
LOCATION: (164)...(871)
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(1060)
OTHER INFORMATION: n = A,T,C or G
US-09-910-151-1
```

Query Match 99.6%; Score 862.8; DB 10; Length 1060;  
Best Local Similarity 99.8%; Pred. No. 3,6e-277;  
Matches 864; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1 GGCACCTACTCCCGAGCTAAGGAGGAGAGCTGATCACCATGAATATGCTTCTATT 60
Db 123 GACACTTACTCCCGAGCTAAGGAGGAGAGCTGATCACCATGAATATGCTTCTATT 182
Qy 61 TGGGTGCTCTGCTGGAGCATTTTCTTTGCTGATCTGATCTGTTCAAGAGAGACCTTG 120
Db 183 TGGGTGCTCTGCTGGAGCATTTTCTTTGCTGATCTGATCTGTTCAAGAGAGACCTTG 242
Qy 121 CTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGTGTGGGGTCTCTATCAAC 180
Db 243 CTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGTGTGGGGTCTCTATCAAC 302
Qy 181 CCAGCTGGGTGTGGGCCCCAGCTCACTGCTATTATTAACCAATCTGAAGTGTGCTGGAA 240
Db 303 CCAGCTGGGTGTGGGCCCCAGCTCACTGCTATTATTAACCAATCTGAAGTGTGCTGGAA 362
Qy 241 ATTTCAAGAGAGAGTGAAGAGCGGTACTGAACAGACAAATTAACCCATTCAGATCGTCC 300
Db 363 ATTTCAAGAGAGAGTGAAGAGCGGTACTGAACAGACAAATTAACCCATTCAGATCGTCC 422
Qy 301 GCTACTGGAAGTCAAGTCAATAGGCCCAAGAGTGAAGTCACTGCTATCAAGCTGGCTA 360
Db 423 GCTACTGGAAGTCAAGTCAATAGGCCCAAGAGTGAAGTCACTGCTATCAAGCTGGCTA 482
Qy 421 CAGGACGTGCTGTCTACTCTCAGGTTTGAAGTGAAGCAAGAAAGAGTGGCCGAGACC 480
Db 543 CAGGACGTGCTGTCTACTCTCAGGTTTGAAGTGAAGCAAGAAAGAGTGGCCGAGACC 602
Qy 541 AACAAAGAAAAAGCCACAGAAATTCCTTATGTGAATTTGTGAAGTATTCAGCCGAA 600
Db 663 AACAAAGAAAAAGCCACAGAAATTCCTTATGTGAATTTGTGAAGTATTCAGCCGAA 722
Qy 601 TTTTGGGAGAGTGGCGGCTTGTCTACTGTCACTGCAAAAGACAGCTCAGGGAATCGAG 660
Db 723 TTTTGGGAGAGTGGCGGCTTGTCTACTGTCACTGCAAAAGACAGCTCAGGGAATCGAG 782
Qy 661 TGGGGCACTTCAATGAGGAGGAGCGTGGCATCTACCAATGTTTAAATATGTATCT 720
```



Db 783 TGGGGCACTTCATGAGGGGAGCGTCGGCATCTACACCAATGTTTAAATATGATCTCT 842  
Qy 721 GGATTGAGAACACTGCTAAGGACAAAGTGAAGACCCCTACTTCCCTGCAATTCACCTGGC 780  
Db 843 GGATTGAGAACACTGCTAAGGACAAAGTGAAGACCCCTACTTCCCTGCAATTCACCTGGC 902  
Qy 781 TCTGCCATGGACTATACAGAGCAATATTTTCCCTCTATTCATTAATTAATTCCTCAATG 840  
Db 903 TCTGCCATGGACTATACAGAGCAATATTTTCCCTCTATTCATTAATTAATTCCTCAATG 962  
Qy 841 AAAATTTGGGAATGTAGCATCTACTAGT 866  
Db 963 AAAATTTGGGAATGTAGCATCTACTAGT 988

RESULT 3  
US-10-246-006-1  
; Sequence 1, Application US/10246006  
; Publication No. US20030144493A1  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell C.  
; TITLE OF INVENTION: Human Serine Protease  
; FILE REFERENCES: 00-16  
; CURRENT APPLICATION NUMBER: US/10/246,006  
; NUMBER OF SEQ. ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ. ID NO 1  
; LENGTH: 865  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (42)...(746)  
US-10-246-006-1

Query Match 98.2%; Score 850.8; DB 12; Length 865;  
Best Local Similarity 99.7%; Pred. No. 3.3e-273;  
Matches 863; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 GGCACTTACTCTCCCTGAGCTTAAGGGGGAAGAGCTGATCAACCAATATATGCTCTTCTATT 60  
Db 1 GGCACTTACTCTCCCTGAGCTTAAGGGGGAAGAGCTGATCAACCAATATATGCTCTTCTATT 60  
Qy 61 TGGGTGCTCTGCTGGGACATTTTCTTTGCTGACTCATCTGTTCAAGAAAGACCCCTG 120  
Db 61 TGGGTGCTCTGCTGGGACATTTTCTTTGCTGACTCATCTGTTCAAGAAAGACCCCTG 120  
Qy 121 CTCCCTATTTGGTGTACTCAAGCTCACTCACTCAACCCCTGTGTGGGGTCTCTCATCAAC 180  
Db 121 CTCCCTATTTGGTGTACTCAAGCTCACTCACTCAACCCCTGTGTGGGGTCTCTCATCAAC 180  
Qy 181 CCAGCTGGGTGCTGGCCCGCAAGCTCACTCACTTATTAACCAATCTGAAGTGTCTGGGAA 240  
Db 181 CCAGCTGGGTGCTGGCCCGCAAGCTCACTCACTTATTAACCAATCTGAAGTGTCTGGGAA 240  
Qy 241 ATTTCAGAGAGAGGTCAAGAGCGTACTGAACAGCAATTAACCCCATTCAGTCTGCTCC 300  
Db 241 ATTTCAGAGAGAGGTCAAGAGCGTACTGAACAGCAATTAACCCCATTCAGTCTGCTCC 300  
Qy 301 GCTACTGGAAGTCAAGTCAATGAGGCCCAAGAGTGAAGTCACTCATCAAGCTGAGCTA 360  
Db 301 GCTACTGGAAGTCAAGTCAATGAGGCCCAAGAGTGAAGTCACTCATCAAGCTGAGCTA 360  
Qy 361 AGCTGTCATGCTCAATTCACCAAGTCAAGCCCTTCCCTGCGACCAACCAATGTCAGGC 420  
Db 361 AGCTGTCATGCTCAATTCACCAAGTCAAGCCCTTCCCTGCGACCAACCAATGTCAGGC 420  
Qy 421 CAGGCACTGTTGTCTACTCTCAAGTTTGAAGTGAAGCAAGAAAGAGTGGCGACACC 480  
Db 421 CAGGCACTGTTGTCTACTCTCAAGTTTGAAGTGAAGCAAGAAAGAGTGGCGACACC 480  
Qy 481 CTGACTTGCGGCAAGACCTGAGAGGCCCGCGTATGTCTGATGGAATGCCAAAAACAG 540

Db 481 CTGACTTGCGGCAAGACCTGAGAGGCCCGCGTATGTCTGATGGAATGCCAAAAACAG 540  
Qy 541 AACAGAGAAAACCCACAGAAATTCCTATATGTTGAAATTTGGAAAGTATTCAGCCGAA 600  
Db 541 AACAGAGAAAACCCACAGAAATTCCTATATGTTGAAATTTGGAAAGTATTCAGCCGAA 600  
Qy 601 TTTTGGGAGAGTGGCCGTTGCTACTGTCATCTGCAAGACAGCTCCAGGAATCGAGG 660  
Db 601 TTTTGGGAGAGTGGCCGTTGCTACTGTCATCTGCAAGACAGCTCCAGGAATCGAGG 660  
Qy 661 TGGGGCACTTCATGGAAGGGAGCGTCGGCATCTACACCAATGTTTACAAATATGATCTCT 720  
Db 661 TGGGGCACTTCATGGAAGGGAGCGTCGGCATCTACACCAATGTTTACAAATATGATCTCT 720  
Qy 721 GGATTGAGAACACTGCTAAGGACAAAGTGAAGACCCCTACTTCCCTGCAATTCACCTGGC 780  
Db 721 GGATTGAGAACACTGCTAAGGACAAAGTGAAGACCCCTACTTCCCTGCAATTCACCTGGC 780  
Qy 781 TCTGCCATGGACTATACAGAGCAATATTTTCCCTCTATTCATTAATTAATTCCTCAATG 840  
Db 781 TCTGCCATGGACTATACAGAGCAATATTTTCCCTCTATTCATTAATTAATTCCTCAATG 839  
Qy 841 AAAATTTGGGAATGTAGCATCTACTAGT 866  
Db 840 AAAATTTGGGAATGTAGCAATCTAGT 865

RESULT 4  
US-09-842-758-11  
; Sequence 11, Application US/09842758  
; Publication No. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Vermet, Corine A. M.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Shinkels, Richard A.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Zehnusen, Bryan D.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Paturajan, Meera  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Gangoli, Saha A.  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Macdougall, John R.  
; APPLICANT: Taupier, Raymond J.  
; APPLICANT: Grose, William M.  
; APPLICANT: Edward, Szeeres S.  
; APPLICANT: Alsbrook II, John P.  
; TITLE OF INVENTION: No. US20030083244A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCES: 15966-783  
; CURRENT APPLICATION NUMBER: US/09/842,758  
; PRIOR APPLICATION NUMBER: 2001-04-25  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: 60/200,158  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,613  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,780  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/201,006  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,007  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,236  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,238  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,186

PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 11  
LENGTH: 730  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-842-758-11

Query Match 84.3%; Score 730; DB 11; Length 730;  
Best Local Similarity 100.0%; Pred. No. 7e-233;

Matches 730; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 GATCACCATGAATATGCTCTATTTGGTGTCTGCTGCGGACATTTTCTTTGCTGA 94  
DB 1 GATCACCATGAATATGCTCTATTTGGTGTCTGCTGCGGACATTTTCTTTGCTGA 60  
QY 95 CTGATCTGTTGAGAAGAACCTGCTCCCTATTGTGTACCTCAAGTCTCACTTCAA 154  
DB 61 CTGATCTGTTGAGAAGAACCTGCTCCCTATTGTGTACCTCAAGTCTCACTTCAA 120  
QY 155 CCCCCTGTGGGCGTCTCTATCAAAACCAAGTGGGTGTGGCCCCAGTCTGCTATT 214  
DB 121 CCCCCTGTGGGCGTCTCTATCAAAACCAAGTGGGTGTGGCCCCAGTCTGCTATT 180  
QY 215 ACCAAATCTGAAGTGTGCTGGGAATTTCAAGAGCAGATCAAGAGCGGTACTGACA 274  
DB 181 ACCAAATCTGAAGTGTGCTGGGAATTTCAAGAGCAGATCAAGAGCGGTACTGACA 240  
QY 275 GACAATTAACCCCATTCAGATCGTCCGCTACTGGAATCAAGTCAATAGCCGCCACAGA 334  
DB 241 GACAATTAACCCCATTCAGATCGTCCGCTACTGGAATCAAGTCAATAGCCGCCACAGA 300  
QY 335 TGACCTCATGCTCATCAAGCTGGGCTAAGCTGCTCAATCCCAAGTCCAGCCCT 394  
DB 301 TGACCTCATGCTCATCAAGCTGGGCTAAGCTGCTCAATCCCAAGTCCAGCCCT 360  
QY 395 TCCCTGCGCCACGCAATGTCAGGCGCAGGCACTGTCTACTCTCAGGTTTGAAGT 454  
DB 361 TCCCTGCGCCACGCAATGTCAGGCGCAGGCACTGTCTACTCTCAGGTTTGAAGT 420  
QY 455 GAGCGCAAGAAAAGTGGCCGACACCTGACTTGGCGCAACCTGAGGCCGCCGCTGAT 514  
DB 421 GAGCGCAAGAAAAGTGGCCGACACCTGACTTGGCGCAACCTGAGGCCGCCGCTGAT 480  
QY 515 GTCTGATCGAATATGCAAGAAAAGCAAGAAAAGCAAGAAAAGCAAGAAAAGCAAG 574  
DB 481 GTCTGATCGAATATGCAAGAAAAGCAAGAAAAGCAAGAAAAGCAAGAAAAGCAAG 540  
QY 575 GAAATTTGTGAAGATTCAGCGCAATTTTGGGAGGTGGCCGTTGCTACTGTCATCTG 634  
DB 541 GAAATTTGTGAAGATTCAGCGCAATTTTGGGAGGTGGCCGTTGCTACTGTCATCTG 600  
QY 635 CAAAGACAAGCTCCAGGGAATCGAGTGGGCACTTCATGAGAGGGGAGCTCGGCACTTA 694  
DB 601 CAAAGACAAGCTCCAGGGAATCGAGTGGGCACTTCATGAGAGGGGAGCTCGGCACTTA 660  
QY 695 CACCAATGTTTCAAAATATGATCTCTGATTTGAAGAACCTGCTAAGGACAAAGTGAACCC 754  
DB 661 CACCAATGTTTCAAAATATGATCTCTGATTTGAAGAACCTGCTAAGGACAAAGTGAACCC 720

QY 755 TACTTCTCCC 764  
DB 721 TACTTCTCCC 730

## RESULT 5

US-09-842-758-13  
Sequence 13, Application US/09842758  
Publication No. US20030083244A1  
GENERAL INFORMATION:  
APPLICANT: Verneet, Corine A. M.  
APPLICANT: Fernandes, Elma R.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Shinkels, Richard A.  
APPLICANT: Malyankar, Uriel M.  
APPLICANT: Boldog, Ferenc L.  
APPLICANT: Zernusen, Bryan D.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Gangoli, Saha A.  
APPLICANT: Smithson, Glenda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R.  
APPLICANT: Taupier, Raymond J.  
APPLICANT: Grose, William M.  
APPLICANT: Edward, Szekeres S.  
TITLE OF INVENTION: No. US20030083244A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT APPLICATION NUMBER: US/09/842,758  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,238  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 730  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-842-758-13

Query Match 84.1%; Score 728.4; DB 11; Length 730;  
Best Local Similarity 99.9%; Pred. No. 2.4e-232;



```
APPLICANT: Burgess, Catherine E
APPLICANT: Gangolli, Esna A
APPLICANT: Smithson, Glenda
APPLICANT: Raetelli, Luca
APPLICANT: MacDougall, John R
APPLICANT: Taulier, Raymond J
APPLICANT: Grose, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsbrook II, John P
TITLE OF INVENTION: No. US20030083244A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 15966-783
CURRENT APPLICATION NUMBER: US/09/842,758
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/200,158
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/200,613
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,780
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/201,006
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,007
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,236
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,238
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,186
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: 60/201,474
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/201,508
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/220,591
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/232,678
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/263,217
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 60/265,160
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 106
LENGTH: 721
TYPE: DNA
ORGANISM: Homo sapiens
US-09-842-758-106

Query Match      81.3%; Score 703.8; DB 11; Length 721;
Best Local Similarity 99.6%; Pred. No. 46-224; Indels 1; Gaps 1;
Matches 716; Conservative 0; Mismatches 2;
```

```
QY 339 CTCATGCTCATCAAGCTGCTAAGCCCTGCATGCTCAATCCCAAGTCCAGCCCTTCCC 398
DB 302 CTCATGCTCATCAAGCTGCTAAGCCCTGCATGCTCAATCCCAAGTCCAGCCCTTACC 361
QY 399 CTGCCACCAACATGTCAGGCCAGGCACTGTCTGTACTTCTTCAGTTTGGACTGGAGC 458
DB 362 CTCGCCACCAACATGTCAGGCCAGGCACTGTCTGTACTTCTTCAGTTTGGACTGGAGC 421
QY 459 CAAGAAAAGAGTGGCCGACACCCCTGACTGCGGAGAACCTGGAGGCCCGGTGATGCT 518
DB 422 CAAGAAAAGAGTGGCCGACACCCCTGACTGCGGAGAACCTGGAGGCCCGGTGATGCT 481
QY 519 GATGAGAAATGCC-AAAAAACAAGAAAAGAAAAGCCACAGAAATCTTATGTGAA 577
DB 482 GATGAGAAATGCCAAAAAACAAGAAAAGAAAAGCCACAGAAATCTTATGTGAA 541
QY 578 ATTGTGAAAGTATTCAGCCGAATTTTGGGAGGTGGCCGTCTCTCTCATCTGCA 637
DB 542 ATTGTGAAAGTATTCAGCCGAATTTTGGGAGGTGGCCGTCTCTCTCATCTGCA 601
QY 638 AGACAGCTCCAGGGAATTCAGGTGGGCACTTCATGAGAGGAGCTGGCATCTACAC 697
DB 602 AGACAGCTCCAGGGAATTCAGGTGGGCACTTCATGAGAGGAGCTGGCATCTACAC 661
QY 698 CAATGTTTACAAATATGATCTGATGATGAGAACCTGCTAAGGACAAGTGAGACCTTA 756
DB 662 CAATGTTTACAAATATGATCTGATGATGAGAACCTGCTAAGGACAAGTGAGACCTTA 720
```

# RESULT 8

US-09-888-615-43

Sequence 43, Application US/09888615

Patent No. US20020064856A1

GENERAL INFORMATION:

APPLICANT: PLOMAN, GREGORY

APPLICANT: WHYTE, DAVID

APPLICANT: CAENEPEL, SEAN

APPLICANT: CHARYDCZAK, GLEN

APPLICANT: MANNING, GERARD

APPLICANT: SUDARSANAM, SUCHA

TITLE OF INVENTION: NOVEL PROTEASES

FILE REFERENCE: 038602/1214

CURRENT APPLICATION NUMBER: US/09/888,615

CURRENT FILING DATE: 2001-06-26

PRIOR APPLICATION NUMBER: 60/214,047

PRIOR FILING DATE: 2000-06-26

NUMBER OF SEQ ID NOS: 150

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 43

LENGTH: 798

TYPE: DNA

ORGANISM: Homo sapiens

US-09-888-615-43

```
Query Match      70.0%; Score 606.4; DB 9; Length 798;
Best Local Similarity 88.6%; Pred. No. 1.56-191;
Matches 707; Conservative 0; Mismatches 1; Indels 90; Gaps 1;

QY 42 ATGAATATATGCTTCTATTTGGGTGTCCTGCTGGGACATTTTCTTGGTCACTCATCT 101
DB 1 ATGAATATATGCTTCTATTTGGGTGTCCTGCTGGGACATTTTCTTGGTCACTCATCT 60
QY 102 GTTCAGAAAGAACCCGCTCCGATTTTGGTGTACCTCAAGTCTCACTTCAACCCCTGT 161
DB 61 GTTCAGAAAGAACCCGCTCCGATTTTGGTGTACCTCAAGTCTCACTTCAACCCCTGT 120
QY 162 GTGGCGTCTCATCAAAACCCAGCTGGGTGCTGGCCCGGCTCACTGATTTTCAAAAT 221
DB 121 GTGGCGTCTCATCAAAACCCAGCTGGGTGCTGGCCCGGCTCACTGATTTTCAAAAT 180
QY 222 CTGAAAGTATGCTGGGAAATTTTCAAGAGCAAGTCAAGACGGTACTGAAACAAT 281
DB 181 CTGAAAGTATGCTGGGAAATTTTCAAGAGCAAGTCAAGACGGTACTGAAACAAT 240
```

Oy	282	AACCCATTCAAGTTCGCCGCTAAGAACTACAGTCAATAGGCGCCACAGATGACTC	3 411
Db	241	AACCCATTCAAGTTCGCCGCTAAGAACTACAGTCAATAGGCGCCACAGATGACTC	3 000
Oy	342	ATGCTCATCAAGCTGACTAAGCTTCGCATGCTCAATCCCAAGTCAAGCCCTTC	4 011
Db	301	ATGCTCATCAAGCTGACTAAGCTTCGCATGCTCAATCCCAAGTCAAGCCCTTC	3 600
Oy	402	GCCACCAACCAATGTCAGGCCAGGCACTGTCTCTACTCTCAGGTTTGGACTGACCA	4 611
Db	361	GCCACCAACCAATGTCAGGCCAGGCACTGTCTCTACTCTCAGGTTTGGACTGACCA	4 200
Oy	462	GAAAAAGT-----	4 707
Db	421	GAAAAAGTGGCTTTGGCACTGAGCCACCAAGGCATCTGACTGCAAGAGCCCA	4 800
Oy	471	-----GGCCGACCTCTGACTGCG	4 911
Db	481	GCCATTCTGATTGGCAGAGACCAATTCATCATGAAACAAGGCCGACCTGACTGCG	5 100
Oy	492	CAGAACCTGGAGGCCCCCGTATGTCTGATCGAAGATGCCAAAAACAGAACAGAAAA	5 511
Db	541	CAGAACCTGGAGGCCCCCGTATGTCTGATCGAAGATGCCAAAAACAGAACAGAAAA	6 000
Oy	552	AGCACACGAATTCCTTATGTGTGAATTGTGAAGATTCAGGCCAAATTTTGGGAG	6 111
Db	601	AGCACACGAATTCCTTATGTGTGAATTGTGAAGATTCAGGCCAAATTTTGGGAG	6 600
Oy	612	GTGGCGTGTCTACTGTCTCATCTGGCAAGAACAAGCTCCAGGGAATGAGGTGGGCACTTC	6 711
Db	661	GTGGCGTGTCTACTGTCTCATCTGGCAAGAACAAGCTCCAGGGAATGAGGTGGGCACTTC	7 200
Oy	672	ATGGAGAGGGAAGTGGCAATCTAACCAATGTTTAAATATGTAATCTGGATTGAGAAC	7 211
Db	721	ATGGAGAGGGAAGTGGCAATCTAACCAATGTTTAAATATGTAATCTGGATTGAGAAC	7 800
Oy	732	ACTGCTTAAGGACAAGTGA	7 49
Db	781	ACTGCTTAAGGACAAGTGA	7 98

```

RESULT 9
US-10-246-006-3
Sequence 3, Application US/10246006
Publication No. US2003014493A1
GENERAL INFORMATION:
APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: Human Serine Protease
FILE REFERENCE: 00-16
CURRENT APPLICATION NUMBER: US/10/246,006
CURRENT FILING DATE: 2002-09-17
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 705
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: This degenerate nucleotide sequence encodes the
OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
FEATURES:
NAME/KEY: misc feature
LOCATION: (1)..(705)
OTHER INFORMATION: n = A,T,C or G
US-10-246-006-3

```

```

Query Match      59.0%; Score 511.2; DB 12; Length 705;
Query Similarity 59.3%; Pred. No. 9.3e-160;
Best Local Similarity 163; Mismatches 124; Indels 0; Gaps 0;
Matches 418; Conservative 163;
42 ATGAATATGCTTATTTGGGTCGCCGCTGGACATTTTCTTGGCTGACACACT 101
|||||:|:| |:|:|:| |:| |:| |:| |:| |:| |:| |:|:|:|:|

```

Db	1	ATGAARATATGNTTATATATYATYNGNNGTNTYNGCNGNACNTTATTTATTTGCGNATYMSMSN	60
Qy	102	GTTCAAGAAAGAAAGACCCCTGCTCCCTATTTGGTGTACTCAAGTCTCACTTCAACCCCTGT	161
Db	61	GTNCARARARGAAGAACGNCNCNCNTAAYTNGNTATYATYTAAMSMNCATYTTAAVCCNTG	120
Qy	162	GTGGGCGCTCATCAATAAACCCAGCTGGGTGCTGGGCCCGCAGTCACTGCTATTATTAACCAAT	221
Db	121	GTNGNGNTNTATHTBARCCNMKSNSTGGGNTYNTGNCNCNGCNCATYGTATATYNTCCMAA	180
Qy	222	CTGAAATGATGCTGGGAAATTTCAAGACGAGTCAAGACGGTACTGAACAGACAATT	281
Db	181	YTNARGTNATAGYTNAGNAAATTTAARWSNMNGNTMGATGAGNACNBARCARPACNATH	240
Qy	282	AAACCCATTCAGATGTCGCGCTACTGGAACCTACAGTCATATAGGGCCCCACAGAGTGACTC	341
Db	241	AAATCCNATHCARATHTGTMTGNTATYTGAAATATATYMSNCATYMSGNCNCNBARATYATY	300
Qy	342	ATGTCATCAAGCTGGCTAAGCTGAGCTGCATGCTCAATGCCAAGTCCAGCCCTTCCCTC	401
Db	301	ATGTATNATHAARATYTGCAARCCNCCNATGYTAAATCCAAAGTNCARCCNTATNACNTY	360
Qy	402	GCCACCAACCATNTGTCAGGCGCAAGGCACTGTCTGTCACTCAAGTTTGACATCGAGACCA	461
Db	361	GCNACNACNAAATGTMNGCNCNGNACNGNTNGYATNTYNTMSNGNTNATATATYGGWNCAR	420
Qy	462	GAATAACAGTGGCCGACACCTGACTCTTGCGGCGAGAACCTGAGAGCCCCCTGTATGTCGAT	521
Db	421	GABAAATYMSNGMNGNCAYTCNGAAYTNTGNCABAAATYTGARGCNCNCNTNATGWSNGAY	480
Qy	522	CGAAGATTCGCAAAAAACAGAACAGAAAGAAAAACCAAGAAATTCCTTATGTGTGAATTT	581
Db	481	MNGNARTGYCARAAACNBARCARGANAARWSNCAATYMGAAATWSNTNTNGYTNAAARTY	540
Qy	582	GTGAAGATATTCAGCGCAATTTTGTGGGAGAGGCGCGTGTACTGTCACTGTCGAAGAC	641
Db	541	GTNAARGTNTTYTWSMNGAATHTTGTGNCBARGTNGCNGTNGCAACNGTNAATHTHTAARAT	600
Qy	642	AACTCCACAGGAATTCAGAGTGGGCACTTCATGGAAGGGGACGTGGCATCTAACCAAT	701
Db	601	AARATYTCNARGNATHGARGTNGNCATYTTATYAGGNGNGGATGTINGNATHTATYACNAAY	660
Qy	702	GTTTACAAATATGATCCGTGATTTAGAAACACTGCTAAGACAAG	746
Db	661	GTNTATATATATYATYNTMSNTGATHTBARAAATYACNCGMAARGAATAAAR	705

```

RESULT 10
US-10-101-510-199
: Sequence 199. Application US/10101510
: Publication No. US20030148295A1
: GENERAL INFORMATION:
: APPLICANT: WAN, JACKSON
: APPLICANT: WANG, YIXIN
: TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
: FILE REFERENCE: 15117, 0012
: CURRENT APPLICATION NUMBER: US/10/101,510
: CURRENT FILING DATE: 2002-03-20
: PRIOR APPLICATION NUMBER: 60/276,947
: PRIOR FILING DATE: 2001-03-20
: NUMBER OF SEQ ID NOS: 805
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 199
: LENGTH: 853
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-101-510-199

```

Query Match	17.8%;	Score 154.4;	DB 12;	Length 853;
Best Local Similarity	55.0%;	Pred. No. 1.9e-40;		
Matches 349;	Conservative 0;	Mismatches 276;	Indels 9;	Gaps 2.
109 AAGAAGACCCGCTCCCTATTGCTGACCTCAAGTCACTTCAACCCCTGTGTGGCG 168				

Db 135 AGGAGAAATTCCTCCCTACAGAGGATGCTCCGAAATTCCTGCTCCCACTTTCGCGTGGCT 194  
Qy 169 TCTCTATCAAAACCAAGCTGGGTGCTGGCCCAAGCTCACTGCTATTTACCAATCTGAAG 228  
Db 195 CCCCTATCAGGAAACAGTGGGTGATCAGAGCTCACTGTCTAAAGACCCGCAATCCAG 254  
Qy 229 TGATGCTGGGAAATTTCAAGAGCAGAGTCAGAGACGGTACTGAACAGCAATTAACCCCA 288  
Db 255 TGAAGCTGGAGAGACACAACTCAAGTCTGGAGGGAAATGAGCAATTCATCAATGGG 314  
Qy 289 TTCAGATGCTCCGCTACTGAACTAAGTCAAGTCAAGGCCCCAAGAGTCACTGCTCA 348  
Db 315 CCAAGATCATCCGCAACCTTAATACACAGAGCACTCTGAGCAATGATCATGCTGA 374  
Qy 349 TCAAGCTAGCTAAGCTTCCATGCTCAATCCCAAGTCCAGCCCTTCCCTCCGCAACA 408  
Db 375 TCAAACTCTCTCACTCCGCTCAATCAATGCCCCGCTGTCACCAATCTCTGCCCCAG 434  
Qy 409 CCAATGTCAGGCAAGGCACTGCTGCTACTCTCAAGTTTGAATGAGCCAGAAACA 468  
Db 435 CCCCTCAGCTGCTGGCACTGAGTCTCTATCTCCGCTGGAGCAACTCTGAGCTTG 494  
Qy 469 GTGGCCGACACCCCTGACTTGGCGGAGAACTGAGGCCCCGCTGATGCTGATCGAGAT 528  
Db 495 GTGCTGACTACCCAGACAGAGCTGAAGTCTGAGATGCTCCGCTGTCACCCAGGCTGAGT 554  
Qy 529 GCCAAAAACAGAACAGAAAGAACCAAGAGATCTTATGCTGAAATTTGTGAAG 588  
Db 555 GTAAAGCTCTCTACCTGGAAGATTAACCAAGCATGTTCTGTGTGGCTTCTCTGAGG 614  
Qy 589 TATTCAGCCGAATTTTGGGAGAGTGGCGCTTCTACTGTCATCTGCAAGAGCAAGCTTC 648  
Db 615 GAGGAGAGATTTCTGCGAGGCTGACTTGTGGCTCTGTGCTCTGCAACGAGCTTC 674  
Qy 649 AGGGAATCG--AGGTGGGGCACTTTCATGAGGAGGAGCTC-----GGCATCTACACCA 699  
Db 675 AAGAGTGTCTCCCTGGGGGCAATGCTGTGCTGGAAGAAACAGAGCTGAGTCAACA 734  
Qy 700 ATGTTTACAAATATGATCTGATCTGATGAGAACAC 733  
Db 735 AGTCTACAACTATGTGAGCTGATTAAGACAC 768

RESULT 11  
US-09-954-456-748  
Sequence 748, Application US/09954456  
Patent No. US20020115057A1  
GENERAL INFORMATION:  
APPLICANT: Young, Paul  
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can  
FILE OF INVENTION: Sets  
FILE REFERENCE: 689290-76  
CURRENT APPLICATION NUMBER: US/09/954,456  
CURRENT FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US/60/233,617  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/234,052  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,923  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,134  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,637  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,638  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,711  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,720  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,840  
PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,863  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 2276  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 748  
LENGTH: 850  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-954-456-748

Query Match 17.8%; Score 154; DB 10; Length 850;  
Best Local Similarity 55.1%; Pred. No. 2.6e-40; Indels 9; Gaps 2;  
Matches 348; Conservative 0; Mismatches 275; CntrCCAGCC

Qy 111 GAAGACCTGCTCCCTATTTGTTGTTACTAGTCTCACTTCAACCCCTGTGGGCTC 170  
Db 134 GAGAAATCTCTCCCTACAGAGTGTCTCTGAATTTGTGGCTCCACTTGTGGGTGCC 193  
Qy 171 CTCATCAACCCAGCTGGGTGCTGGCCCAAGCTCACTGTAATTTACCAATCTGAAGTG 230  
Db 194 CTCATCAAGCGAACAGTGGGTGATATCAGACAGTCACTGTAACAAGACCCCATCAGGTG 253  
Qy 231 ATGCTGGAAATTTCAAGAGCAGAGTCAGAGCGTACTGAACAGACAATTAACCCATT 290  
Db 254 AGACTGGAGAGCAACATCAAGTCTGAGGGGAATAGCATTTTCAATCAATGCGGCC 313  
Qy 291 CAGATGTCGCTACTGTAATCACTAGTCAATAGGCCCCAGATGACTCATGCTCATC 350  
Db 314 AAGATCACTCCGACCTTAATACACAGGACACTGTGACAAATGATCATCTGATC 373  
Qy 351 AAGCTGCTAAGCTGCTCATGCTCAATCCAAAGTCAAGCCCTTCCCTGCGACCAAC 410  
Db 374 AAATCTCTCACTGCGCTGATCAATGCGCGGTGTCACATCTCTGCGCACCGCC 433  
Qy 411 AATGTCAGGCGACGCTGCTGTCTACTCTCAAGTTTGAAGTGAAGCAAGAAACAGT 470  
Db 434 CCTCAGCTGTGCACTGAGTGTCTCATCTCCGCTGGGCAACACTGAGCTTTGT 493  
Qy 471 GCGCGACACCCCTGACTTGGCGGAGAACCTGAGGCCCCCGTGTGATCTGATCGAAATGC 530  
Db 494 GCTGACTACCCAGACAGCTGAAGTCTGTGATCTGATCTGATCGAGCTGAGTGT 553  
Qy 531 CAAAAACAGAACAGAAAGAACAGAGATTTCTTATGTGAAATTTGAAAGTA 590  
Db 554 AAGGCTCTTACCTCGGAAGATTAACACAGCATGTTGTGTGGCTTCTGAGGGA 613  
Qy 591 TTGACCGGAATTTTGGGAGAGTGGCGGTGCTACTGCTCATCTGCAAGAACACTCCAG 650  
Db 614 GCGAAGGATTCCTGCGAGCGTGAATCTGTGTGGCCCTGTGTCTCAACGAGACCTCCAA 673  
Qy 651 GGAATCG--AGTGGGGCACTTCATGGAAGGAGAGCTC-----GGCATCTACACCAAT 701  
Db 674 GAGGTGTCTCTCGGGGCAATGCTGTGCTGGAAGAACAGGCTGAGTCTACACCAAG 733  
Qy 702 GTTTCAAAATATGATCTGATCTGATGAGAACAC 733  
Db 734 GTTACAACTATGTGAGCTGATTAAGACAC 765

RESULT 12  
US-09-923-779-148  
Sequence 148, Application US/09923779  
Patent No. US20020076721A1  
GENERAL INFORMATION:  
APPLICANT: Pyle, Ruth A.  
APPLICANT: Xu, Jiangshun  
APPLICANT: Kalos, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121-553  
CURRENT APPLICATION NUMBER: US/09/923,779  
CURRENT FILING DATE: 2001-08-06  
NUMBER OF SEQ ID NOS: 155



SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 148  
LENGTH: 802  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-923-779-148

Query Match 17.7%; Score 153.2; DB 9; Length 802;  
Best Local Similarity 53.9%; Pred. No. 4.6e-40;  
Matches 342; Conservative 0; Mismatches 283; Indels 9; Gaps 1;

109 AAGAAGACCTGCTCCCTATTGTTGTTACCTCAAGTCTCACTTCAACCCCTGTGTGGCG 168  
DB AGAGAAATTCTGTCCTTACAGAGTCTTGAATTCGTGCTACCACTTCTGGCGTCT 157  
QY 169 TCCTCATCAAAACCAAGTGGGTGTGCGCCCAAGTCTGCTATTACCAATCTGAAG 228  
DB 158 CCTCATCAGCAAGTGGGTGTGCGCCCAAGTCTGCTATTACCAATCTGAAG 217  
QY 229 TGATGCTGGGAAATTTCAAGAGCAAGTCAAGACGGTACTGAACAGCAATTAACCCA 288  
DB 218 TGAGACTGTGAGAGCAACATCGAAGTCTGAGAGGGGAATGAACAGTTCAATGCGG 277  
QY 289 TTCAAGTCTGCTGCTACTGGAATCAAGTCAAGCCGCCCAAGATGACCTCACTCA 348  
DB 278 CCAAGATCATCCGCAACCAATATCAACAGCGGACTTGGAACATGACATCTGCTGA 337  
QY 349 TCAAGTGGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTGCGCACCA 408  
DB 338 TCAAGTCTGCTGCTGCTGCTCAATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 397  
QY 409 CCAATGCTAGGCGCAGCACTGTCTGTCTACTCTCAGGTTTGAGTGGAGCAAGAAAA 468  
DB 398 CCCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457  
QY 469 GTGGCGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528  
DB 458 GTGCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 517  
QY 529 GCCAAAAACAGAAACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 588  
DB 518 GTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 577  
QY 589 TATTGACCGCAATTTTGGGAGGTGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 648  
DB 578 GAGGCAAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 637  
QY 649 AGGGAATCGAGTGGGCACTTCATGAGGAGGAGCGTGGCA-----TCTACACCA 699  
DB 638 AAGGAATTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 697  
QY 700 ATGTTTCAAAATATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 733  
DB 698 AGGTCTACCACTATGTGACTGATTAAGGACAC 731

RESULT 13  
US-10-101-510-210  
Sequence 210, Application US/10101510  
Publication No. US20030148295A1  
GENERAL INFORMATION:  
APPLICANT: MAN, JACKSON  
APPLICANT: WANG, YIXIN  
TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE  
FILE REFERENCE: 15117.0012  
CURRENT APPLICATION NUMBER: US/10/101.510  
PRIOR FILING DATE: 2002-03-20  
PRIOR APPLICATION NUMBER: 60/276,947  
NUMBER OF SEQ ID NOS: 805  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 210  
LENGTH: 802

TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-101-510-210

Query Match 17.7%; Score 153.2; DB 12; Length 802;  
Best Local Similarity 53.9%; Pred. No. 4.6e-40;  
Matches 342; Conservative 0; Mismatches 283; Indels 9; Gaps 1;

109 AAGAAGACCTGCTCCCTATTGTTGTTACCTCAAGTCTCACTTCAACCCCTGTGTGGCG 168  
DB AGAGAAATTCTGTCCTTACAGAGTCTTGAATTCGTGCTACCACTTCTGGCGTCT 157  
QY 169 TCCTCATCAAAACCAAGTGGGTGTGCGCCCAAGTCTGCTATTACCAATCTGAAG 228  
DB 158 CCTCATCAGCAAGTGGGTGTGCGCCCAAGTCTGCTATTACCAATCTGAAG 217  
QY 229 TGATGCTGGGAAATTTCAAGAGCAAGTCAAGACGGTACTGAACAGCAATTAACCCA 288  
DB 218 TGAGACTGTGAGAGCAACATCGAAGTCTGAGAGGGGAATGAACAGTTCAATGCGG 277  
QY 289 TTCAAGTCTGCTGCTACTGGAATCAAGTCAAGCCGCCCAAGATGACCTCACTCA 348  
DB 278 CCAAGATCATCCGCAACCAATATCAACAGCGGACTTGGAACATGACATCTGCTGA 337  
QY 349 TCAAGTGGCTAAGCTGCTGCTCAATCCCAAGTCCAGCCCTTCCCTGCGCACCA 408  
DB 338 TCAAGTCTGCTGCTGCTGCTCAATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 397  
QY 409 CCAATGCTAGGCGCAGCACTGTCTGTCTACTCTCAGGTTTGAGTGGAGCAAGAAAA 468  
DB 398 CCCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457  
QY 469 GTGGCGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528  
DB 458 GTGCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 517  
QY 529 GCCAAAAACAGAAACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 588  
DB 518 GTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 577  
QY 589 TATTGACCGCAATTTTGGGAGGTGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 648  
DB 578 GAGGCAAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 637  
QY 649 AGGGAATCGAGTGGGCACTTCATGAGGAGGAGCGTGGCA-----TCTACACCA 699  
DB 638 AAGGAATTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 697  
QY 700 ATGTTTCAAAATATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 733  
DB 698 AGGTCTACCACTATGTGACTGATTAAGGACAC 731

RESULT 14  
US-10-181-808-1  
Sequence 1, Application US/10181808  
Publication No. US20030157634A1  
GENERAL INFORMATION:  
APPLICANT: Polymun Scientific Immunobiologische Forschung GmbH  
APPLICANT: Matanovich, Diethard  
APPLICANT: Kättinger, Hermann  
APPLICANT: Hohenblum, Hubertus  
APPLICANT: Nascherberger, Stefan  
TITLE OF INVENTION: METHOD FOR THE MANUFACTURE OF RECOMBINANT TRYPSIN  
FILE REFERENCE: 2235-146  
CURRENT APPLICATION NUMBER: US/10/181.808  
PRIOR FILING DATE: 2002-07-23  
PRIOR APPLICATION NUMBER: PCT/EP01/00770  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: US 60/177,348  
NUMBER OF SEQ ID NOS: 8



SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 1  
 LENGTH: 699  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-181-808-1

Query Match 16.4%; Score 142; DB 12; Length 699;  
 Best Local Similarity 52.8%; Pred. No. 2,3e-36;  
 Matches 335; Conservative 0; Mismatches 290; Indels 9; Gaps 1;

109 AAGAAGACCTGCTCCCTATTGGTGTACCTCAAGTCTCACTTCAACCCCTGTGTGGCG 168  
 47 AGAGAAATTTCTGCTCCCTACAGGTGCTCCGAAATTTGGCTACATTTCTGTGTGGCT 106  
 169 TCTCATCAAAACCAAGCTGGGTGCTGGCCCAAGCTCACTGCTATTATTCAAATCTGAAG 228  
 107 CCTCATCAACGAACAGTGGGTGTATCAGAGGCCCACTGCTCAAGTCCCGCATCCAGG 166  
 229 TGATGCTGGGAAATTTCAAGACAGTCAAGACGGTACTGAAACAGACAAATTAAACCCA 288  
 167 TGAGACTGGGAGACAAACATGCAAGTCTGAGGGGAAATGACAGTTCAATCAATCAG 226  
 289 TTGAGATGCTCCCTACTGTAGACCTACAGTCAATAGCGCCCAAGAGATGACCTCATGCTCA 348  
 227 CCAAGATCATCCGCAACCCCAATACAGACAGAAAGTCTGAAACATGACATCTGTTAA 286  
 349 TCAAGCTGGCTAAAGCTGCTCAATCCCAAGTCTCAAGCCCTTCCCTCGCCACA 408  
 287 TCAAGCTCTCTCTGCTGCTGATCAACGCGCGGTGTCACATCTCTGCGCCACG 346  
 409 CCAATGTCAGGCGCAGGACTGTCTGTACTCTCAAGTTTGGACTGAGCCAAAGAAACA 468  
 347 CCCCTCAGCCACTGGGACGAGAGTGCCTCACTCTGGCTGGGCAACAATCGAGCTCTG 406  
 469 GTGGCCGACACCTGACTTGGGAGGAAACCTGGAGGCCCGCGTATGTCTGATGAGAT 528  
 407 GCGCGGACTACCGAGAGAGTGCAGTGCCTGTGATGCTCTGTCTGAGCCAGCTAAGT 466  
 529 GCCAAACAAACAGAACAGAAAGCAACAGAAATTCCTTATGTGTGAATTTGTGAAG 588  
 467 GTGAAGCTCTCTACCTCGGAAGATTAACAGCAACATTTCTGTGTGGCTTCTTGAAG 526  
 589 TATTCAGCCGAATTTTGGGAGGTGCGCTGTCTACTGTCTATCTGCAAGACAAAGTCC 648  
 527 GAGGCAAGATTCATGTCAAGGTATTTCTGTGGCCCTGTGTCTGCAATGAGACAGCTCC 586  
 649 AGGAAATCGAG-----GTGGGCACTTCATGAGGAGGAGAGTGGCATCTACACA 699  
 587 AAGAGATGTCTCTCTGGGGTATGTGCTGCGCCAGAGAACAAAGCTGAGTCTACACA 646  
 700 ATGTTACAATATGATCTGTGATTTGAGAACAC 733  
 647 AGGTCTACACTACGTGAATGATTAAGAACAC 680

## RESULT 15

US-09-923-779-91  
 Sequence 91, Application US/09923779  
 Patent No. US20020076721A1  
 GENERAL INFORMATION:  
 APPLICANT: Pyle, Ruth A.  
 APPLICANT: Xu, Jiangchun  
 APPLICANT: Kalos, Michael D.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 FILE REFERENCE: 210121.553  
 CURRENT APPLICATION NUMBER: US/09/923,779  
 CURRENT FILING DATE: 2001-08-06  
 NUMBER OF SEQ ID NOS: 155  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 91  
 LENGTH: 522

TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 3, 5  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-923-779-91

Query Match 15.3%; Score 132.8; DB 9; Length 522;  
 Best Local Similarity 55.1%; Pred. No. 2,3e-33;  
 Matches 260; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

109 AAGAAGACCTGCTCCCTATTGGTGTACCTCAAGTCTCACTTCAACCCCTGTGTGGCG 168  
 50 AGAGAAATTTCTGCTCCCTACAGGTGCTCCGAAATTTGGCTACATTTCTGTGTGGCT 109  
 169 TCTCATCAAAACCAAGCTGGGTGCTGGCCCAAGCTCACTGCTATTATTCAAATCTGAAG 228  
 110 CCTCATCAACGAACAGTGGGTGTATCAGAGGCCCACTGCTCAAGTCCCGCATCCAGG 169  
 229 TGATGCTGGGAAATTTCAAGACAGTCAAGACGGTACTGAAACAGACAAATTAAACCCA 288  
 170 TGAGACTGGGAGACAAACATGCAAGTCTGAGGGGAAATGACAGTTCAATCAATCAG 229  
 289 TTGAGATGCTCCCTACTGTAGACCTACAGTCAATAGCGCCCAAGAGATGACCTCATGCTCA 348  
 230 CCAAGATCATCCGCAACCCCAATACAGACAGCGGATCTGGAACATGACATCTGCTGA 289  
 349 TCAAGCTGGCTAAAGCTGCTCAATCCCAAGTCTCAAGCCCTTCCCTCGCCACA 408  
 290 TCAAGCTCTCTCTGCTGCTGATCAACGCGGTGTCACATCTCTGCGCCACG 349  
 409 CCAATGTCAGGCGCAGGACTGTCTGTACTCTCAAGTTTGGACTGAGCCAAAGAAACA 468  
 350 CCCCTCAGCTGCTGGGACGAGTGCCTCACTCTGGCTGGGCAACAATCGAGCTCTG 409  
 469 GTGGCCGACACCTGACTTGGGAGGAAACCTGGAGGCCCGCGTATGTCTGATGAGAT 528  
 410 GTGGCGACTACCGAGAGAGTGCAGTGCCTGTGATGCTCTGTCTGAGCCAGCTGAGT 469  
 529 GCCAAACAAACAGAACAGAAAGCAACAGAAATTCCTTATGTGTGAATTTGTGAAG 580  
 470 GTGAAGCTCTCTACCTCGGAAGATTAACAGCAACATTTCTGTGTGGCTT 521

Search completed: November 8, 2003, 00:24:39  
 Job time : 357 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 8, 2003, 01:23:17 ; Search time 29 Seconds  
(without alignments)  
5743.592 Million cell updates/sec

Title: US-10-037-270-482  
Perfect score: 1560  
Sequence: 1 ggcactactccctgagcta.....tgggaatgtagcatcactagt 866

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 56616

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xjh  
-Q=/cgn2\_1/USPRO.spool/US10037270/runat\_07112003\_140516\_8805/app.query.fasta\_1.1031  
-DB=PIR 76 -QMT=faetan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10037270.qcgn\_1.1\_25@runat\_07112003\_140516\_8805 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	377.5	24.2	248	2	trypsin (EC 3.4.21)
2	353	22.6	304	2	trypsin (EC 3.4.21)
3	352.5	22.6	247	1	trypsin (EC 3.4.21)
4	351.5	22.5	246	2	trypsin (EC 3.4.21)
5	350.5	22.5	246	1	trypsin (EC 3.4.21)
6	350.5	22.5	259	2	trypsin (EC 3.4.21)
7	349	22.4	243	2	trypsin (EC 3.4.21)
8	346.5	22.2	246	1	trypsin (EC 3.4.21)
9	346	22.2	247	1	trypsin (EC 3.4.21)
10	346	22.2	247	2	trypsin (EC 3.4.21)
11	344.5	22.1	248	2	trypsin (EC 3.4.21)
12	340.5	21.8	246	1	trypsin (EC 3.4.21)
13	335.5	21.5	231	1	trypsin (EC 3.4.21)
14	334.5	21.4	247	1	trypsin (EC 3.4.21)

15	329.5	21.1	247	2	trypsin (EC 3.4.21)
16	328.5	21.1	247	2	trypsin (EC 3.4.21)
17	312.5	20.0	229	1	trypsin (EC 3.4.21)
18	312	20.0	229	1	trypsin (EC 3.4.21)
19	312	20.0	247	2	trypsin (EC 3.4.21)
20	311.5	20.0	228	2	trypsin (EC 3.4.21)
21	311	19.9	246	2	trypsin (EC 3.4.21)
22	309	19.8	246	2	trypsin (EC 3.4.21)
23	287	18.4	242	2	trypsin (EC 3.4.21)
24	285	18.3	242	2	trypsin (EC 3.4.21)
25	280	17.9	242	2	trypsin (EC 3.4.21)
26	280	17.9	242	2	trypsin (EC 3.4.21)
27	264.5	17.0	240	2	trypsin (EC 3.4.21)
28	260	16.7	241	2	trypsin (EC 3.4.21)
29	247.5	15.9	255	2	trypsin (EC 3.4.21)
30	244.5	15.7	250	2	trypsin (EC 3.4.21)
31	244	15.6	248	2	trypsin (EC 3.4.21)
32	243	15.6	260	2	trypsin (EC 3.4.21)
33	237	15.2	248	2	trypsin (EC 3.4.21)
34	235.5	15.1	261	2	trypsin (EC 3.4.21)
35	235	15.1	248	2	trypsin (EC 3.4.21)
36	231	14.8	248	2	trypsin (EC 3.4.21)
37	228.5	14.6	233	1	trypsin (EC 3.4.21)
38	228.5	14.6	247	1	trypsin (EC 3.4.21)
39	221	14.2	261	1	trypsin (EC 3.4.21)
40	216.5	13.9	812	1	trypsin (EC 3.4.21)
41	215	13.8	246	2	trypsin (EC 3.4.21)
42	214	13.7	253	2	trypsin (EC 3.4.21)
43	214	13.7	261	2	trypsin (EC 3.4.21)
44	209.5	13.4	263	1	trypsin (EC 3.4.21)
45	208.5	13.4	250	2	trypsin (EC 3.4.21)

#### ALIGNMENTS

RESULT 1  
S55067  
trypsin (EC 3.4.21.4) I precursor, pancreatic - chicken  
N/Alternate names: trypsinogen I  
C/Species: Gallus gallus (chicken)  
C/Date: 28-Oct-1996 #sequence, revision 07-Feb-1997 #text\_change 21-Jul-2000  
C/Accession: S55067; S72345; S55065; S72346; S71155  
R/Wang, K.; Gan, L.; Lee, I.; Hood, L.  
Biochem. J. 307, 471-479, 1995  
A/Title: Isolation and characterization of the chicken trypsinogen gene family.  
A/Reference number: S55065; M01D:95251611; PMID:7733885  
A/Accession: S55067  
A/Molecule type: mRNA  
A/Residues: 1-248 <MAN1>  
A/Cross-references: EMBL:U15156; NID:G603904  
A/Experimental source: clone 1-P38  
A/Accession: S72345  
A/Molecule type: DNA  
A/Residues: 1-248 <MAN2>  
A/Cross-references: EMBL:U15155; NID:G603902; PIDN:AAA79912.1; PID:G603903  
A/Experimental source: clone 1-P38  
A/Accession: S55065  
A/Molecule type: mRNA  
A/Residues: 1-9, 'V', '11-12', 'T', '14-102', 'A', '104-214', 'I', '216-248' <MAN3>  
A/Cross-references: EMBL:U15155; NID:G603902; PIDN:AAA79912.1; PID:G603903  
A/Experimental source: clone 1-P38  
A/Accession: S72346  
A/Molecule type: DNA  
A/Residues: 1-9, 'V', '11-12', 'T', '14-102', 'A', '104-214', 'I', '216-248' <MAN4>  
A/Cross-references: GB:U15155; NID:G603902; PIDN:AAA79912.1; PID:G603903  
A/Experimental source: clone 1-P38  
R/Wang, K.  
submitted to the EMBL Data Library, September 1994  
A/Reference number: S71155  
A/Accession: S71155  
A/Molecule type: mRNA  
A/Residues: 1-102, 'A', '104-248' <MAN5>  
A/Cross-references: EMBL:U15156; NID:G603904; PIDN:AAA79913.1; PID:G603905  
A/Experimental source: clone 1-P38

C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-25/Domain: activation peptide #status predicted <APT>  
F:26-246/Product: trypsin I #status predicted <MNT>  
F:26-241/Domain: trypsin homology <TRY>  
F:65,109,202/Active site: His, Asp, Ser #status predicted

## Alignment Scores:

Pred. No.:	7.79e-26	Length:	248
Score:	377.50	Matches:	88
Percent Similarity:	53.06%	Conservative:	42
Best Local Similarity:	35.92%	Mismatches:	100
Query Match:	24.20%	Indels:	15
DB:	2	Gaps:	4

US-10-037-270-482 (1-866) x S55067 (1-248)

```
Oy 42 ATGAATAT-----GTCTCTATTGGGTGCTCGTGGACATTTTCTTGT 92
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 1 MetlyserleuValleuValAlaPheleuGlyVal---AlaValAlaPheProIleSer 19

Oy 93 GACTCATCTGTTCAGAA-----GAAGACCTCTCTCTAT 128
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 20 AspGluAspAspAspLysIleValGlyGlyTyrSerCysAlaArgSerAlaAlaProTyr 39

Oy 129 TTGGTGTACCTCAAGTCTCATCTCAACCCCTGTGGCGCTTCATCAACCAAGCTGG 188
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 40 GluValSerLeuAsnSerGlyTyrHisPheCysGlyGlySerLeuIleSerSerGlnTyr 59

Oy 189 GTGGTGGCCCGCCATCTGCTATTTACCAATCTGAAGTGAAGTGAAGTGAAGTTCAG 248
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 60 ValLeuSerAlaAlaHisCysTyrLysSerSerIleGlnValLysLeuGlyGlnTyrAsn 79

Oy 249 AGCAGAGTCAAGACCGTACTGAAACAGACAATTAACCCATTCAGATCGTCCGTACTGG 308
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 80 LeuAlaIaGlnAspGlySerGlnGlnThrIleSerSerSerValIleArgHisSer 99

Oy 309 AACATACATCTAGCGCCACAGAGATGACTCATGCTCATCAAGTGGCTTAAGCTGCC 368
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 100 GlyTyrAsnSerAsnThrLeuAsnAspIleMetLeuIleLysLeuSerLysAlaAla 119

Oy 369 ATGCTCATCCCAAGTCCAGCCCTTCCCTCGCCACCAATGTCAGGCCAGGCACT 428
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 120 ThrLeuAsnSerTyrValAsnThrValProLeuProThrSerCysValThrAlaGlyThr 139

Oy 429 GTCTGTACTCTGAGTTTGAGTGAAGTGAAGCAAGAAACAGTGGCCGACCTGACTTG 488
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 140 ThrCysLeuIleSerGlyTyrGlyAsnThrLeuSerSerGlySerLeuTyrProAspVal 159

Oy 489 CGGCAGAACCTGAGAGCCCGCTGATGCTGATGAGAAATGCCAAAAACAGAAACAGA 548
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 160 LeuGlnCysLeuAsnAlaProValLeuSerSerSerGlnCysSerSerAlaTyrProGly 179

Oy 549 AAAAGCCACAGAGATTCCTATGCTGAATTTTGGAAGTATTGACCGCAATTTTGGG 608
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 180 ArgIleThrSerAsnMetIleCysAlleGlyTyrLeuAsnGlyLysAspSerCysGln 199

Oy 609 GAGGTGGCCGTGTACTGTGATCTGATGCAAGACAGTCCAGGGAATC-----GAG 659
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 200 GlyAspSerGlyGlyProValValCysAsnGlyGlnLeuGlnGlyPheValSerTyrGly 219

Oy 660 GTGGGCGACTTCATGGAGGAGGAGCTCGGCATCTACACCAATGTTTAAATATGATCC 719
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 220 IleGlyCysAlaGlnLysGlyTyrProGlyValTyrThrLysValCysAsnTyrValSer 239

Oy 720 TGGATTGAGAACACT 734
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 240 TrpIleLysThrThr 244
```

RESULT 2  
S33496  
trypsin (EC 3.4.21.4) IV form a - human

C:Species: Homo sapiens (man)  
C:Date: 03-Mar-1994 #sequence\_revision 03-Aug-1995 #text\_change 15-Aug-1997  
C:Accession: S33496  
R:Legend, U.: Corbach, S.; Mann, A.; Kang, J.; Mueller-Hill, B.  
Submitted to the EMBL Data Library, March 1993  
A:Description: Identification, cloning and characterization of a cDNA encoding a human B.

A:Reference number: S33496  
A:Accession: S33496  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-304 <MNT>  
A:Cross-references: EMBL:X12781  
C:Genetics:

A:Gene: GDB:PRSS4; TRY4  
A:Cross-references: GDB:335300  
A:Map position: 7q35-7q35  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:81-296/Domain: trypsin homology <TRY>  
F:120,164,257/Active site: His, Asp, Ser #status predicted

## Alignment Scores:

Pred. No.:	1.28e-23	Length:	304
Score:	353.00	Matches:	87
Percent Similarity:	51.38%	Conservative:	43
Best Local Similarity:	34.39%	Mismatches:	113
Query Match:	22.63%	Indels:	10
DB:	2	Gaps:	4

US-10-037-270-482 (1-866) x S33496 (1-304)

```
Oy 6 TTATCCCTGAGCTAAGGGGGAAGAGCTGATGATCAATCAATATATGTC-----TTC 56
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 47 LeuHisProLeuLeuGlyGlyArgThrTyrArgAlaAlaArgAspAlaAspGlyCysGlu 66

Oy 57 TATTTGGGTGTCTCGCTGGGACATTTTCTTCTGATCATCTGTTCAG----- 107
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 67 AlaLeuGlyThrValAlaValAlaProPheAspAspAspLysIleValGlyGlyTyrThr 86

Oy 108 ---AAAGAACCTGCTCCCTATTTGCTACTCTCAAGTCTCAACCTGCTG 164
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 87 CysGlnGlnAsnSerLeuProTyrGlnValSerLeuAsnSerGlySerHisPheCysGly 106

Oy 165 GGGCTCCATCAAAACCCAGCTGGGCTGGCCCGCCAGCTCACTGCTATTTACCAATCTG 224
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 107 GlySerLeuIleSerGlnGlnThrValValSerAlaAlaHisCysTyrLysThrArgIle 126

Oy 225 AAAGTATGCTGGGAAATTTCAAGACAGAGTCAAGACGCTATGAAACAGCAATTPAC 284
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 127 GlnValArgLeuGlyGlnHisAsnIleLysValLeuGlnGlyAsnGlnPheIleAsn 146

Oy 285 CCCATTCAATGCTCGCTGCTACTGAGAACTACAGTATAGCCGCCACAGATGACTCATG 344
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 147 AlaAlaLysIleIleArgHisProLysTyrAsnArgAspThrLeuAspAsnAspIleMet 166

Oy 345 CTCATCAAGTGGGTAGCGCTGATGCTGATGCTCAATCCAAAGTCCAGCCCTTCCCTGCC 404
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 167 LeuIleLysLeuSerSerProAlaValIleAsnAlaArgValSerThrIleSerLeuPro 186

Oy 405 ACCACCAATGTCAGGCCAGGACATGTGTCTACTCTCAGATTTGACCTGAGCAAGAA 464
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 187 ThrIaProProAlaAlaGlyThrGlnCysLeuIleSerGlyTyrGlyAsnThrLeuSer 206

Oy 465 AACAGTGGCCGACACCTGACTGCTGGCAGAACTGAGAGCCCGCTGATGCTGATGCA 524
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 207 PheGlyAlaAspTyrProAspGlnLeuLysCysLeuAspAlaProValLeuThrGlnAla 226

Oy 525 GAATGCCAAAAAAGCAAGAAAGCAAGAAAGCAAGAAATCTATGAGTGAATTTGTG 584
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 227 GluCysLysAlaSerTyrProGlyLysIleThrAsnSerMetPheCysValGlyPheLeu 246

Oy 585 AAAGTATTCAGCCGAATTTTGGGAGAGTGGCCGTGCTACTGATCTGCAAGAACAG 644
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
```

Db 247 GluGlyGlyLysAspSerCysGlnArgAspSerGlyGlyProValValCysAsnGlyGln 266  
 QY 645 CTCAGGAGG---ATCGAGTGGGGCACTTCATGGGA-----GGGAGCGTGGCATCTAC 695  
 Db 267 LeuGlnIleValValSerTrpGlyIleGlyCysAlaTrpLysAsnArgProGlyValTyr 286  
 QY 696 ACCAATGTTTACAAATATGATCTCGATTGAGAACACT 734  
 Db 287 ThrLysValTyrAsnTyrValAspTrpIleLysAspThr 299

RESULT 3  
 A25852  
 trypsin (EC 3.4.21.4) I precursor [validated] - human  
 N/Alternate names: trypsin, catenonic; trypsinogen I  
 C/Species: Homo sapiens (man)  
 C/Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 08-Dec-2000  
 C/Accession: A25852; B61066; A43988  
 R/Entl, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.  
 Gene 41, 305-310, 1986  
 A/Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human  
 A/Reference number: A91544; MUID:86221712; PMID:3011602  
 A/Accession: A25852  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <EMI>  
 A/Cross-references: GB:M22612; NID:9521215; PIDN:AAA61231.1; PID:9521216  
 R/Kimland, M.; Russick, C.; Marks, W.H.; Borgstrom, A.  
 Clin. Chim. Acta 184, 31-46, 1989  
 A/Title: Immunoreactive anionic and cationic trypsin in human serum.  
 A/Reference number: A61066; MUID:90091010; PMID:2598466  
 A/Accession: B61066  
 A/Molecule type: protein  
 A/Residues: 16-43 <KIM>  
 R/Kolunen, E.; Huhtala, M.L.; Stenman, U.H.  
 J. Biol. Chem. 264, 14095-14099, 1989  
 A/Title: Human ovarian tumor-associated trypsin. Its purification and characterization  
 A/Reference number: A43988; MUID:89340515; PMID:2503510  
 A/Accession: A43988  
 A/Molecule type: protein  
 A/Residues: 16-54 <KOI>  
 A/Experimental source: mucinous ovarian tumor cyst fluid  
 C/Genetics:  
 A/Gene: GDB:PRSS1; TRY1  
 A/Cross-references: GDB:119620; OMIM:276000  
 A/Map position: 7q35-7q35  
 A/Note: The human genome contains at least ten trypsin genes or pseudogenes, at least tw  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; pancreas; phosphoprotein; protein digestion; serine proteinase; z  
 F/1-15/Domain: signal sequence #status predicted <SIG>  
 F/16-246/Product: trypsinogen I #status experimental <ZYM>  
 F/16-23/Domain: activation peptide #status predicted <ENZ>  
 F/24-246/Product: trypsin I #status predicted <APT>  
 F/24-239/Domain: trypsin homology <TRY>  
 F/30-160/48-64,139-206,171-185,196-220/Disulfide bonds: #status predicted  
 F/63,107,200/Active site: His, Asp, Ser #status predicted  
 F/75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Alignment Scores:  
 Pred. No.: 1.4e-23 Length: 247  
 Score: 352.50 Matches: 82  
 Percent Similarity: 50.43% Conservative: 36  
 Best Local Similarity: 35.04% Mismatches: 109  
 Query Match: 22.60% Indels: 7  
 Gaps: 2

US-10-037-270-482 (1-866) x A25852 (1-247)

QY 54 TTCATTGGGTGTCGTCGGAGACATTTTCTTGCTAGCATGTTAG----- 107  
 Db 9 PheValAlaAlaAlaValAlaAlaProPheAspAspAspLysIleValGlyGlyTyr 28  
 QY 108 -----AAAGAGACCTGCTCCTTATTTGGTGTACTCAAGTCTCAACCCCTGT 161  
 Db 29 AsnGlyGluGluAsnSerValProTyrGlnValSerLeuAsnSerGlyTyrHisAspCys 48

QY 162 GTGGGCTCTCATCAACCCAGCTGGGTCTGGCCCAAGCTCATCTTATACCAAT 221  
 Db 49 GlyGlySerLeuIleAsnGlnGlnTrpValValSerAlaGlyIleCysTyrLysSerArg 68  
 QY 222 CTGAAGTGTGCTGGGAAATTTCAAGCAGATCAGACCGTACTGAACCAATT 281  
 Db 69 IleGlnValArgLeuGlyGlnHisAsnIleGlnValLeuGlnGlyAsnGlnGlnPhe 88  
 QY 282 AACCCATTCAGATCGTCCGCTACTGGAATCAAGTCATAGCCGCCACAGATGACTC 341  
 Db 89 AsnAlaAlaLysIleIleArgHisProGlnTyrAspArgLysThrLeuAsnAspIle 108  
 QY 342 ATGCTCATCAAGCTGGCTGAAGCTGCATGCTCAATCCCAAGTCCAGCCCTTCCCTC 401  
 Db 109 MetLeuIleLysLeuSerSerArgAlaValIleAsnAlaArgValSerThrIleSerLeu 128  
 QY 402 GCCACCAACCAATGTACAGCCAGCCACTGTCTGTCTACTCTCAGCTTGAAGCCAA 461  
 Db 129 ProThrAlaProProAlaThrGlyThrLysCysLeuIleSerGlyTyrGlyAsnThrAla 148  
 QY 462 GAAACAGTGGCCGACACCTGACTTGGCAGAACCTGAGGCCCCCGTGAATGTGAT 521  
 Db 149 SerSerGlyAlaAspTyrProAspGluLeuGlnCysLeuAspAlaProValLeuSerGln 168  
 QY 522 CGAATGCGCAAAACAGAACAGAAAGCCACAGAAATTCCTATGTGAAATTT 581  
 Db 169 AluGlyCysGluLysSerTyrProGlyLysIleThrSerAsnMetPheCysValGlyPhe 188  
 QY 582 GTGAAGTATTCACCGCAATTTTGGGAGGTGGCGGTGCTACTGTCATCTGCCAAGAC 641  
 Db 189 LeuGlnGlyGlyLysAspSerCysGlnGlyLysSerGlyGlyProValValCysAsnGly 208  
 QY 642 AACCTCCAGGATCCAGGTC-----GGGACTTCAATGGAGGAGCGTGGCATC 692  
 Db 209 GlnLeuGlnGlyValValSerTrpGlyAspGlyCysAlaGlnLysAsnLysProGlyVal 228  
 QY 693 TACACCAATGTTACAAATATGATCTGATCTGATGAGAACACT 734  
 Db 229 TyrThrLysValTyrAsnTyrValLysTrpIleLysAsnThr 242

RESULT 4  
 B25528  
 trypsin (EC 3.4.21.4) precursor - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 22-Jun-1999  
 C/Accession: B25528  
 R/Stevenson, B.J.; Hagenbuechle, O.; Wellauer, P.K.  
 Nucleic Acids Res. 14, 8307-8330, 1986  
 A/Title: Sequence organization and transcriptional regulation of the mouse elastase II ar  
 A/Reference number: A93646; MUID:87066713; PMID:3641189  
 A/Accession: B25528  
 A/Molecule type: mRNA  
 A/Residues: 1-246 <STE>  
 A/Cross-references: GB:X04574; NID:954918; PIDN:CAA28243.1; PID:954919  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: calcium binding; hydrolase; protein digestion; serine proteinase  
 F/1-23/Domain: signal sequence #status predicted <SIG>  
 F/24-246/Product: trypsin #status predicted <MAT>  
 F/30-160/48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
 F/63,107,200/Active site: His, Asp, Ser #status predicted  
 F/75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Alignment Scores:  
 Pred. No.: 1.72e-23 Length: 246  
 Score: 351.50 Matches: 77  
 Percent Similarity: 53.49% Conservative: 38  
 Best Local Similarity: 35.81% Mismatches: 97  
 Query Match: 22.53% Indels: 3  
 Gaps: 1

US-10-037-270-482 (1-866) x B25528 (1-246)



A/Cross-references: EMBL:X1345; NID:9405755  
 C/Genetics:  
 A/Intons: 25/3; 78/1; 162/3; 208/2  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; serine proteinase  
 F;37-251/Domain: trypsin homology <TRY>  
 F;75,119,212/Active site: His, Asp, Ser #status predicted

## Alignment Scores:

Pred. No.:	2,13e-23	Length:	259
Score:	350.50	Matches:	87
Percent Similarity:	51.1%	Conservative:	42
Best Local Similarity:	34.52%	Mismatches:	114
Query Match:	22,47%	Indels:	9
DB:	2	Gaps:	4

US-10-037-270-482 (1-866) x 138363 (1-255)

```

QY 6 TPACTCCCTGAGTAAAGGGGAAAGCTGGATCCATGAATATGTC-----TTC 56
Db 3 LeuHieProLeuLeuGlyArgThrTrpArgAlaIaIaArgAspAlaAspGlyCysGlu 22
QY 57 TATTTGGGTGCTCCCTGGGACATTTTCTTGGTGAATCTGTT-----CAG 107
Db 23 AlaLeuGlyThrValAlaValProPheAspAspAspAspAspAspAspAspAsp 42
QY 108 AAGAAGACCTCCCTCCCTATTTGGTGTACCTCAAGTCACTTCAACCCCTGTGGGC 167
Db 43 CysGluAspSerLeuProGlyGlnValSerLeuAspSerLeuSerHisPheCysGlyGly 62
QY 168 GTCTCTCCCTCAACCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 227
Db 63 SerLeuLeuSerLeuGlnGlnTrpValValSerAlaIaIaIaIaIaIaIaIaIaIa 82
QY 228 GTGATCTGGGAAATTTCAAGACAGAGTCAAGACGGTCACTGAACAGACAAATTAACCC 287
Db 83 ValArgLeuGlyGlnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 102
QY 288 ATTCAATGCTCCGCTACTGAGACTACAGTCATAGGCCCCCAAGATGACCTCATCTC 347
Db 103 AlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 122
QY 348 ATCAACCTGGCTAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 407
Db 123 IleLeuLeuSerLeuProAlaValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 142
QY 408 ACCAATGTCAGGCGACGCTGTGTCTTCTCTCACTGAGTTTGACTGAGCCCAAGAAC 467
Db 143 AlaProProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 162
QY 468 AGTGGCCGACACCTGACTTGGCGGACACCTGAGGCCCTGGATGTCTGATCGAGAA 527
Db 163 GlnAlaAspArgProAspGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 182
QY 528 TGGCAAAAAACAAGAAAAAGCAAGAAATCTTATGTGTGAAATTTGGAAA 587
Db 183 CysLeuAlaSerLeuProGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 202
QY 588 GTATTCAGCCGAATTTTGGGAGGTGGCGCTGTCTACTGTCTACTGTCAAGCAAGCTC 647
Db 203 GlnGlyLeuAspSerCysGlnArgAspSerGlyGlyProValValCysAsnGlnLeu 222
QY 648 CAGGGA---ATCAGAGTGGGGCACTTCATGGG-----GGGACGTGGCATCTACAC 698
Db 223 GlnGlyValValSerTrpGlnHisGlyCysAlaTrpLeuAsnArgProGlyValTyrThr 242
QY 699 AATGTTTACAATATATCTGTGATTTGAGAACCT 734
Db 243 LysValTyrAsnTyrValAspTrpIleLeuAspTrp 254

```

RESULT 7  
 A35871  
 trypsin (EC 3.4.21.4) precursor, pancreatic - African clawed frog

C/Species: Xenopus laevis (African clawed frog)  
 C/Date: 09-Nov-1990 #sequence\_reviseion 09-Nov-1990 #text\_change 04-Mar-2000  
 C/Accession: A35871; S12117  
 R/Shi, Y.B.; Brown, D.D., 1990  
 Genes Dev. 4, 1107-1113, 1990  
 A/Title: Developmental and thyroid hormone-dependent regulation of pancreatic genes in X.  
 A/Reference number: A35871; PMID:91007255; PMID:2210372  
 A/Accession: A35871  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-243 <SHI>

## A/Cross-references:

A/Cross-references: EMBL:X53458; NID:965162; PIDN:CA17538.1; PID:965163  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; protein digestion; serine proteinase  
 F;15-20/Domain: signal sequence #status predicted <SIG>  
 F;15-20/Domain: activation peptide #status predicted <AP>  
 F;21-236/Domain: trypsin homology <TRY>  
 F;26-243/Product: trypsin I #status predicted <MAT>  
 F;27-157,45-61,129-230,136-203,168-182/Dileuflide bonds: #status predicted  
 F;60,104,197/Active site: His, Asp, Ser #status predicted

## Alignment Scores:

Pred. No.:	2,89e-23	Length:	243
Score:	349.00	Matches:	78
Percent Similarity:	51.88%	Conservative:	46
Best Local Similarity:	32.64%	Mismatches:	107
Query Match:	22,37%	Indels:	8
DB:	2	Gaps:	2

US-10-037-270-482 (1-866) x A35871 (1-243)

```

QY 42 ATGAATATGCTCTTATTTGGTGTCCCTGGGACATTTTCTTGTGATCTATCT 101
Db 1 MetLeuPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 20
QY 102 GTT-----CAGAAAGAACCTGCTCCCTATTTGGTGTGATCTCAAGTCT 146
Db 21 IleIleGlyGlyAlaThrCysAlaLeuSerSerValProTyrIleValSerLeuAsnSer 40
QY 147 CACTTCAACCCCTGTGTGGCGCTGCTCATCAACCCAGCTGGTGTGGTGGTGGTGGT 206
Db 41 GlnTyrHisPheCysGlyGlySerLeuIleThrAsnGlnTrpValValSerAlaIaIa 60
QY 207 TGTATTTACCAATTTGAAAGTGAAGTGTGGGAAATTTCAAGACAGTCAAGACGCT 266
Db 61 CysTyrLeuAlaSerIleGlnValArgLeuGlyGlnHisAsnIleAlaLeuSerGlnGly 80
QY 267 ACTGACAGACAAATTAACCCCATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 326
Db 81 ThrGlnGlnPheIleSerSerSerLeuValIleArgHisSerGlyTyrAsnSerTyrThr 100
QY 327 CCACAGATGACCTCATGCTCATCAAGCTGAGCTGACCTGCTGCTGCTGCTGCTGCTGCT 386
Db 101 LeuAspAsnAspIleWetLeuIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 120
QY 387 CAGCCCTTCCCTCCGACACCAATGTCAGGCCAGCTGTCTGTCTACTCTCAGCT 446
Db 121 AsnThrValProLeuProSerLeuGlyCysSerAlaIaIaGlyThrSerCysLeuIleSerGly 140
QY 447 TTGAGCTGGAGCAAGAAACAGTGGCCGACACCTGATCTGAGCGAGAACTGAGGCC 506
Db 141 TrpGlyAsnThrLeuSerAsnGlnSerAsnTyrProAspLeuLeuGlnCysLeuAsnAla 160
QY 507 CCCGATGCTGTGATGAGAAATGCCAAAAACAAGAAACAAGAAACAAGAAATTC 566
Db 161 ProIleLeuThrAsnAlaGlnCysAsnSerAlaTyrProGlyGlnIleThrAlaAsnMet 180
QY 567 TTTATGTGAAATTTTGTGAAGTATTCACGCCAATTTTGGGAGGTGGCCGTGCTACT 626
Db 181 IleCysValGlyTyrMetGlnGlnGlyLeuAspSerCysGlnGlyAspSerGlyGlyPro 200
QY 627 GTCATCTGCAAGACAGACCTCCAGGAATC-----GAGGTGGGCACTTCATGGGA 677

```



Db 201 ValValCySaSnAGInLeuGInGlyValValSerTrpGlyTyrGlyCySaLeMeArg 220

Qy 678 GGGAGCTGGCATCTACACCAATGTTTACAATATGATCTGGATTGAGACACT 734

Db 221 AsnTyrProGlyValTyrThrIleValCySaSnTyrAsnAlaTyrIleGlnAsnThr 239

RESULT 8

TRRT1

trypsin (EC 3.4.21.4) I precursor - rat

N:Alternate names: trypsinogen I

C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Dec-1982 nosequence\_revision 17-Dec-1982 #text\_change 24-Sep-1999

C:Accession: B22657; A00948

R:Caik, C.S.; Choo, O.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.

J. Biol. Chem. 259, 14255-14264, 1984

A:Title: Structure of two related rat pancreatic trypsin genes.

A:Reference number: A22657; MUID:85054880; PMID:6094547

A:Accession: B22657

A:Molecule type: DNA

A:Residues: 1-246 <CRA>

A:Cross-references: GS:J00778; NID:g206507; PID:AAA98518.1; PID:g206508

A>Note: the authors translated the codon ATC for residue 6 as Ieu and GAC for residue 17

J. Biol. Chem. 257, 9724-9732, 1982

A:Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of

A:Reference number: A00948; MUID:82265624; PMID:6896710

A:Accession: A00948

A:Molecule type: mRNA

A:Residues: 1-246 <MAC>

A:Cross-references: GS:J00778; NID:g206507; PID:AAA98518.1; PID:g206508

C:Genetics:

A:introns: 14/1; 67/2; 152/3; 197/3

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-23/Domain: activation peptide #status predicted <APT>

F:24-246/Product: trypsin I #status predicted <ENZ>

F:24-235/Domain: trypsin homology <IRK>

F:30-160/48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted

F:63,107,200/Active site: His, Asp, Ser #status predicted

F:63,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Alignment Scores:

Pred. No.: 4,866-23 Length: 246

Score: 346.50 Matches: 76

Percent Similarity: 54.50% Conservative: 39

Best Local Similarity: 36.02% Mismatches: 93

Query Match: 22.21% Indels: 3

DB: 1 Gaps: 1

US-10-037-270-482 (1-866) x TRRT1 (1-246)

Qy 111 GAGAGCCTGGCTCTATTTGGTGTACCTCAAGTCTCACTCAACCCCTGTGTGGAGCTC 170

Db 32 GlnHisSerValProTyrGlnValSerLeuAsnSerGlyTyrHisPheCysGlyIleYer 51

Qy 171 CTCATCAACCCAGCTGGGTGGGCCCAAGCTCACTGTTTACCAATCTGAAAGTG 230

Db 52 LeuIleAsnAspGlnTyrValValSerAlaIleHisCysTyrIleSerArgIleVal 71

Qy 231 ATGCTGGAAATTTCAAGACGAGTCAAGACGCTACTGAACAGCAATTAACCCACTT 290

Db 72 ArgLeuGlyGlnHisAsnIleAsnValLeuGlnIleAspGlnGlnPheIleAsnAla 91

Qy 291 CAGATCGTCCGCTACTGGAGTACAGTCACTAGCGGCCCAAGATGAGCTACTGCTATC 350

Db 92 LysIleIleIleYHisProAsnTyrSerSerTyrPheLeuAsnAspIleIleLeuIle 111

Qy 351 AAGCTGCTAAGCTGCGCATGCTCAATCCCAAGTCCAGCCCTTCCCTCGCCACACC 410

Db 112 LysLeuSerSerProValLysLeuAsnAlaArgValAlaProValAlaLeuProSerAla 131

Qy 411 AATGTGAGCGCAGGACACTGTCTCTACTCTCAGGTTTGGACTGAGCCAAAGAAACAGT 470

Db 132 CysalaProAlaGlyThrGlnCysLeuIleSerGlyTrpGlyValanthrThrLeuSerAsnGly 151

Qy 471 GGGCAGACCCCTGACTTCGCGCAGAACTGGAGAGCCCCCGTGAATGTGTGATCGAAGATGC 530

Db 152 ValAsnAnProAspLeuLeuGlnCysValAspAlaProValLeuSerGlnAlaAspCys 171

Qy 531 CAAAAAACGAACAAAGAAAAAAGCCACAGAGAATTCCTTAACTGGGAATTTGTGAAGTA 590

Db 172 GlnAlaAlaIleTrpProGlyIleThrSerSerMetIleCysValGlyPheLeuGlnGly 191

Qy 591 TTCAGCCGAATTTTGGGAGGCGCGTGTGCTACTGTCTCATCTTCGCAAAAGCAAGCTCCAG 650

Db 192 GlyIysAspSerCysGlnGlyAspSerGlyGlyProValValCysAsnGlyLeuGln 211

Qy 651 GGAATC-----GAGGTGGGCACTTCATGAGAGGAGGAGCGTCCGATCTACCAAT 701

Db 212 GlyIleValSerTrpGlyIleThrCysAlaLeuProAspAnProGlyValIleThrIys 231

Qy 702 GTTTCACAAATATGATCCTCGATTCGATTCGAAACACT 734

Db 232 ValCysAsnPheValGlyTrpIleGlnAspThr 242

RESULT 9

TRDG

trypsin (EC 3.4.21.4) precursor, anionic - dog

NAlternate names: cationic trypsinogen

CSpecies: Canis lupus familiaris (dog)

CDate: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 18-Jun-1999

CAccession: A26273

RPinaky, S.D.; LaForge, K.S.; Scheele, G.

Mol. Cell. Biol. 5, 2669-2676, 1985

Article: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequence

AReference number: A26273; MUID:86284628; PMID:3841794

AAccession: A26273

A:Molecule type: mRNA

A:Residues: 1-247 <PIN>

A:Cross-references: GB:MI1589; NID:G164094; PIDN:AAA30899.1; PID:G164095

C:Superfamily: trypsin, trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F.1-15/Domain: signal sequence #status predicted <SIG>

F.16-23/Domain: activation peptide #status predicted <AP>

F.24-247/Product: trypsin, anionic #status predicted <ENZ>

F.24-239/Domain: trypsin homology <TRY>

F.30-160, 48-64, 132-233, 133-208, 171-185/Disulfide bonds: #status predicted

F.63, 107, 200/Active site: His, Asp, Ser #status predicted

F.75, 77, 80, 85/Binding site: calcium (Glu, Asn, Val, Gln) #status predicted

Alignment Scores:

	5,39e-23	Length: 247
Pred. No.:	346.00	Matches: 76
Score:	52.564	Conservative: 47
Percent Similarity:	52.484	Mismatches: 103
Best Local Similarity:	22.184	Indels: 8
Query Match:	1	Gaps: 2

US-10-037-270-482 (1-866) x TRDG (1-247)

Qy 57 TATTGGAGGTGCTCCCTCGTGGACATTTTCTTTCGATCATTCTGTTTCAG----- 107

Db 9 PheLeuGlnAlaIleValAlaIleThrProThrAspAspAspAspIleValGlyIleTrp 28

Qy 108 -----AAGAAGACCCCTGCTCCCTAATTGGTGTACTCTCAAGTCTCACTCAACCCCTGT 161

Db 29 ThrCysGlnGluAsnSerValProGlyGlnValSerLeuAsnAlaGlyTyrHisPheCys 48

Qy 162 GTGGGCGTCTCATCAACCCAGCTGGGTCTGGCCCGCAGTCACTGCTATTACCAAT 221

Db 49 GlyIleSerLeuIleSerAspGlnTrpValValSerAlaAlaHisCysTyrIleSerArg 68

Qy 222 CTGAAAGTATGCTGGGAATTTCAAGACGACAGACAGACGCTATCGAAGACAAT 281

Db 69 IieGlnValArgLeuGlnGlyIleGlyTyrAsnIleAspValLeuGlnGlyAsnGlnGlnPheIle 88



QY 282 AACCCCATCATGATCGTCGGCTACTGGAATCAATGATCAATGAGCCGCCAACGATGACCTC 341  
 Db 89 AmsSerAlaIyValIleIhrghisProAbnityrAsnSerTrpIleuAspAsnIle 1088  
 QY 342 ATGCTCATCAAGCTGGCTTAAGCCTGCATGCTCAATCCCAAGTCGACCCCTTCCCTC 401  
 Db 109 MetIleuIleIyLeuSerSerProAlaValIleuAsnAlaArgAlaIaThrIleSerLeu 1288  
 QY 402 GCCACCAACCAATGTGAGCCGACATGTCGTCTGCTCACTGACAGTTTGACCTGACCCAA 461  
 Db 129 ProAlaGAlaCyAlaAlaIaProGlyIhrGInCySLeuIleSerIlyTrpGlyAsnThrLeu 148  
 QY 462 GAAAAAGTGGCCGACCCCTGATCGGACGACATCTGGAGGCCCCCGATGATGTGAT 521  
 Db 149 SerSerGlyThrAsnTyProGInLeuLeuGInCyLeuAspIlaProIleuThrGIn 168  
 QY 522 CGAGAAATCCCAAAAAACGACAAAGAAAAAGCAAGAAATTCCTTAATGTGAAATTT 581  
 Db 169 AlaGInCySgAlaIaSerTyProGlyIhrGInIleThrGluAsnMetIleCySAlaGlyPhe 188  
 QY 582 GTGAAGAATTCAGCCGAATTTTGGGAGAGTGCCCGTGTCTCTCATCTGCAAGAAG 641  
 Db 189 LeuGInGlyGlyLeaAspSerCySgInGlyAspSerGlyGlyProValIaCySAsnGly 208  
 QY 642 AAGCTCCACGGGAATC-----GAGGTGGGCACTTAAGGAGGGAGCGTGGCATC 692  
 Db 209 GluLeuGInGlyIleValaSerTrpGlyTyArgIyCySAlaGlnIyAsnIySProGlyIVal 228  
 QY 693 TACACCAATGTTTACAAATATGATCTCGATTTAGAACACT 734  
 Db 229 TyrThrIySValCyAsnIyValaSerTrpIleGInSerThr 242

C:Trypsin (EC 3.4.21.4) III precursor - human  
 C/Species: Homo sapiens (man)  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #ext\_change 22-Jun-1993  
 C/Accession: S12764  
 R/Tant, T.; Kawashima, T.; Mita, K.; Taktiguchi, Y.  
 Nucleic Acids Res. 18, 1631, 1990  
 A/Title: Nucleotide sequence of the human pancreatic trypsinogen II cDNA.  
 A/Reference number: S12764; MUID:90221895; PMID:2326201  
 A/Accession: S12764  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <TAN>  
 A/Cross-references: EMBL:X15505; NID:g37459; PIDN:CA33527.1; PID:g37460  
 C/Genetics:  
 A/Gene: GDB:PRSS3; TRY3  
 A/Cross-references: GDB:335297  
 A/Map position: 7q35-7q35  
 C/Superfamily: trypsin  
 C/Keywords: calcium binding; hydrolase; pancreas; protein digestion; serine proteinase  
 F/1-15/Domain: signal sequence #status predicted <Sig>  
 F/16-21/Domain: activation peptide #status predicted <Act>  
 F/22-24//Product: trypsin III #status predicted <Mat>  
 F/24-239/Domain: trypsin homology <TRY>  
 F/30-160,48-64,139-206,171-185/Disulfide bonds: #status predicted  
 F/63,107,200/Active site: His, Asp, Ser #status predicted  
 F/75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Alignment Scores:	
Pred. No.:	5.39e-23
Score:	346.00
Percent Similarity:	52.14%
Best local Similarity:	33.76%
Query Match:	22.18%
DB:	2
Length:	247
Matches:	79
Conservative:	43
Mismatches:	104
Indels:	8
Gaps:	3

```

US-10-037-270-482 (1-866) X SL2764 (1-247)
QY 57 TATTGGGNGTCTCGCTGGACATTTTCCTTGGCTGACTATCTGTGAG-----107
      :::::|||||

```

[illegible]

RESULT 11  
S55066  
trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-P29) - chicken  
N/Alternate names: trypsinogen II  
C/Species: Gallus gallus (chicken)  
C/Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 22-Jun-1999  
C/Accession: S55066; #72347  
R/Wang, K.; Gan, L.; Lee, I.; Hood, L.  
Biochem. J. 307, 471-479, 1995  
A/Title: Isolation and characterization of the chicken trypsinogen gene family  
A/Reference number: S55065; MUID:95251611; PMID:7731885  
A/Accession: S55066  
A/Molecule type: mRNA  
A/Residues: 1-248 <MAN1>  
A/Cross-references: EMBL:U01517; NID:g603906; PIDD:AAA79914.1; PID:g603907  
A/Experimental source: clone 2-P29  
A/Accession: S72347  
A/Molecule type: DNA  
A/Residues: 1-248 <MAN2>  
A/Cross-references: EMBL:U01517; NID:g603906; PIDD:AAA79914.1; PID:g603907  
A/Experimental source: clone 2-P29  
A/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F.1-16/Domain: signal sequence #status predicted <SIG>  
F.17-25/Domain: activation peptide #status predicted <APT>  
F.26-248/Product: trypsin II #status predicted <MAT>  
F.26-241/Domain: trypsin homology <TR>



Db 221 AlaleuProAspAnProGlyValTyrThrIleValCysAsnTyrValAspTrpIleGln 240  
 QY 729 AACACT 734  
 Db 241 AspThr 242

## RESULT 13

TRPSTR  
 trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)  
 N:Contains: trypsinogen  
 C:Species: Sus scrofa domestica (domestic pig)  
 C>Date: 24-Apr-1984 #sequence, revision 24-Apr-1984 #text\_change 31-Mar-2000  
 C/Accession: A90641; A90368; A00947  
 R/Charles, M.; Roversy, M.; Guidoni, A.; Desnuelle, P.  
 Biochim. Biophys. Acta 69, 115-129, 1963  
 A>Title: Su le trypsinogene et la trypsin de porc.  
 A/Reference number: A90641  
 A/Accession: A90641  
 A/Molecule type: protein  
 A/Residues: 1-10 <CHA>  
 R/Hermanson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.  
 Biochemistry 12, 3146-3153, 1973  
 A>Title: Determination of the amino acid sequence of porcine trypsin by sequenator analy  
 A/Reference number: A90368; MUID:73258692; PMID:4738933  
 A/Accession: A90368  
 A/Molecule type: protein  
 A/Residues: 9-231 <HER>  
 A>Note: at position 20, Ile and Val occur alternatively  
 C/Superfamily: trypsin; trypsin homology  
 C/Keyword: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase; zym  
 F:1-231/Product: trypsinogen #status experimental <ZYM>  
 F:1-8/Domain: activation peptide #status experimental <APT>  
 F:9-231/Product: trypsin #status experimental <MAT>  
 F:9-224/Domain: trypsin homology <TRY>  
 F:15-145,33-49,117-218,124-191,156-170,181-205/Disulfide bonds: #status predicted  
 F:48,92,185/Active site: His, Asp, Ser #status predicted  
 F:60,62,65,70/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

## Alignment Scores:

Pred. No.: 4.75e-22 Length: 231  
 Score: 335.50 Matches: 78  
 Percent Similarity: 52.89% Conservative: 41  
 Best Local Similarity: 34.67% Mismatches: 97  
 Query Match: 21.51% Indels: 9  
 DB: 1 Gaps: 2

US-10-037-270-482 (1-866) x TRPSTR (1-231)

QY 69 CTGCTGGAGACATTTTCTTGTGACTGATCTGTTCAGAAAGAACCTGCTCCCTAT 128  
 Db 9 TleValGlyValTyrThrCysAlaAlaAsnSerIle-----ProtYr 22  
 QY 129 TTGCTTACTCAAGTCTCACTTCAACCCCTGTGTGGCGTCCATCAAAACCAAGCTG 188  
 Db 23 GlnValSerLeuAsnSerGlySerHisPheCysGlyIleLeuLeuIleAsnSerGlnTyr 42  
 QY 189 GTTCTGGCCCCCACTCTCTCTATTTTCAAAATCGAAAGTATGCTGGAAATTTCAAG 248  
 Db 43 ValValSerIleAlaHisCysTyrLysSerArgIleGlnValArgLeuGlyGluHisAsn 62  
 QY 249 AGCAGAGTACAGACGGTACTGAAACAGCAATTAAACCCCATTCAGATCGTCCGCTACTGG 308  
 Db 63 IleAspValIleuGluGlyAsnGluGlnPheIleAsnAlaIleValIleIleThrHisPro 82  
 QY 309 AACTACAGTCATAGCCGCCCAACAGATGATCACTGCTCAATCAAGTGTAAAGCTGCC 368  
 Db 83 AsnPheAsnGlyAsnThrLeuAspAsnAspIleMetLeuIleLeuSerSerProAla 102  
 QY 369 ATGCTCAATCCCAAGTCCAGCCCTTCCCTGGCCACCAATGTCAGGCCAGGACT 428  
 Db 103 ThrLeuAsnSerArgValAlaThrValSerLeuProArgSerCysAlaAlaIleGlyThr 122  
 QY 429 GTCTGTACTCTCAGGTTTGACTGAGCCAGCAAGAAACAGTGGCCGACCTGACTTG 488

Db 123 GluCysLeuIleSerGlyTyrGlyAsnThrIleSerSerGlySerSerTyrProSerLeu 142  
 QY 489 CGGCAGACCTTGAGGCCCGCCGATGTCATCGAAGATCCAAAGAAACAGAAACAGGA 548  
 Db 143 LeuGlnCysLeuIleValAlaProValIleuSerAspSerSerCysIleSerTyrProGly 162  
 QY 549 AAAGCCACAGAAATTCCTATGTGTGAAATTTGTGAAAGTATTCAGCCCAATTTTGGG 608  
 Db 163 GlnIleThrGlyAsnMetIleCysValGlyPheLeuGlnIleGlyIleAspSerCysGln 182  
 QY 609 GAGGTGGCCCTTCTACTCTGATCTGCAAGACAGCTCCAGGAAATC-----GAG 659  
 Db 183 GlyAspSerGlyGlyProValValCysAsnGlyGlnLeuGlnIleValSerTyrGly 202  
 QY 660 GTGGGCACTTCAATGAGAGGAGCGTCGATCTACCAATGTTTCAATATGTATCC 719  
 Db 203 TyrGlyCysAlaGlnIleuAsnIleProGlyValTyrThrIleValCysAsnTyrValAsn 222  
 QY 720 TGATTTGAGAACT 734  
 Db 223 TrpIleGlnGlnThr 227

## RESULT 14

B25852  
 trypsin (EC 3.4.21.4) II precursor [validated] - human  
 N:Alternate names: trypsin 2; trypsin, anionic; trypsinogen II  
 C:Species: Homo sapiens (man)  
 C/Date: 03-Mar-1994 #sequence, revision 03-Mar-1994 #text\_change 08-Dec-2000  
 C/Accession: B25852; A61066; B43988  
 R/Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.  
 Gene 41, 305-310, 1986  
 A>Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human I  
 A/Reference number: A91544; MUID:86221712; PMID:3011602  
 A/Accession: B25852  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <EMI>  
 A/Cross-references: GB:M27602; NID:9521217; PIDN:AA61232.1; PID:9521218  
 R/Kimland, M.; Russick, C.; Marke, W.H.; Borgerstroem, A.  
 Clin. Chim. Acta 184, 31-46, 1989  
 A>Title: Immunoreactive anionic and cationic trypsin in human serum.  
 A/Reference number: A61066; MUID:90091010; PMID:2598466  
 A/Accession: A61066  
 A/Molecule type: protein  
 A/Residues: 16-39, 'X', '41-42, 'XXX', '47-49 <KIM>  
 R/Koivunen, E.; Huhkala, M.L.; Stenman, U.H.  
 J. Biol. Chem. 264, 14095-14099, 1989  
 A>Title: Human ovarian tumor-associated trypsin. Its purification and characterization fr  
 A/Reference number: A43988; MUID:89340515; PMID:2503510  
 A/Accession: B43988  
 A/Molecule type: protein  
 A/Residues: 16-49 <KOI>  
 A/Experimental source: mucinous ovarian tumor cyst fluid  
 C/Genetic:  
 A/Genes: GDB:PRS2; TRY2  
 A/Cross-references: GDB:335289; OMIM:601564  
 A/Map position: 7q35-7q35  
 A/Note: the human genome contains at least ten trypsin genes or pseudogenes, at least tw  
 C/Superfamily: trypsin; trypsin homology  
 C/Keyword: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-246/Product: trypsinogen II #status experimental <ENZ>  
 F:16-23/Domain: activation peptide #status experimental <APT>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,171-185,196-220/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

## Alignment Scores:

Pred. No.: 5.88e-22 Length: 247  
 Score: 334.50 Matches: 75  
 Percent Similarity: 52.83% Conservative: 37



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - protein search, using frame\_plus\_n2p model

Run on: November 8, 2003, 00:18:41 ; Search time 17.5 Seconds

(without alignments)  
4654,300 Million cell updates/sec

Title: US-10-037-270-482

Perfect score: 1560

Sequence: 1 ggcactactccctgagctca.....tgggaatgtagcactagc 866

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+n2p,model -DEV=xjh  
-Q=/cgn2\_1/USPRO.spool/US10037270/runat\_07112003\_140514\_8746/app.query.fasta.1.1031  
-DB=SwissProt 41 -QFMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10037270.QCGN\_1\_1\_12@runat\_07112003\_140514\_8746 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-NO TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376.5	24.1	248	1 TRV2_CHICK	Q90628 gallus gall
2	375.5	24.1	248	1 TRV1_CHICK	Q90627 gallus gall
3	353	22.6	304	1 TRV3_HUMAN	P35030 homo sapien
4	352.5	22.6	247	1 TRV1_HUMAN	P07477 homo sapien
5	351.5	22.5	246	1 TRV2_MOUSE	P07146 mus musculu
6	350.5	22.5	246	1 TRV1_CANPA	P06871 canis famli
7	349	22.4	243	1 TRV1_XENLA	P19799 xenopus lae
8	346.5	22.2	246	1 TRV1_RAT	P00762 rattus norv
9	346	22.2	247	1 TRV2_CANPA	P06872 canis famli
10	344.5	22.1	248	1 TRV3_CHICK	Q90629 gallus gall
11	340.5	21.8	246	1 TRV2_RAT	P00761 mus musculu
12	335.5	21.5	231	1 TRV2_MOUSE	P07478 homo sapien
13	334.5	21.4	247	1 TRV2_HUMAN	P70055 xenopus lae
14	331.5	21.2	244	1 TRV2_XENLA	P70055 xenopus lae
15	329.5	21.1	247	1 TRV2_BOVIN	P08426 bos taurus
16	328.5	21.1	247	1 TRV3_RAT	P08426 bos taurus
17	312.5	20.0	243	1 TRV1_BOVIN	P00760 bos taurus
18	312	20.0	229	1 TRV1_SQUAC	P00764 equalus aca

19	312	20.0	247	1 TRV4_RAT	P12788 rattus norv
20	311.5	20.0	238	1 TRV3_SALSA	P35033 salmo salar
21	311	19.9	246	1 TRV3_RAT	P32832 rattus norv
22	309	19.8	246	1 TRV4_RAT	P32831 rattus norv
23	287.5	18.4	244	1 TRV6_HUMAN	Q92876 homo sapien
24	285	18.3	231	1 TRV2_SALSA	P35032 salmo salar
25	280	17.9	242	1 TRV1_SALSA	P35031 salmo salar
26	274.5	17.6	256	1 TRV6_HUMAN	Q92875 homo sapien
27	269.5	17.3	241	1 TRV1_GADMO	P16049 gadus morhu
28	260	16.7	241	1 TRV3_GADMO	Q91041 gadus morhu
29	253	16.2	251	1 TRV6_HUMAN	Q90903 homo sapien
30	252	16.2	260	1 TRV6_HUMAN	Q60259 homo sapien
31	247.5	15.9	255	1 CATG_HUMAN	P08311 homo sapien
32	244	15.6	248	1 GRV1_RAT	Q06605 rattus norv
33	243	15.6	260	1 GRV1_MOUSE	Q61955 mus musculu
34	237	15.2	277	1 NRPN_RAT	Q88780 rattus norv
35	236.5	15.2	277	1 TRV6_HUMAN	Q90903 homo sapien
36	235.5	15.1	261	1 CATG_MOUSE	P28233 mus musculu
37	235	15.1	248	1 GRV6_MOUSE	P08882 mus musculu
38	232	14.9	250	1 TRV6_HUMAN	Q90877 homo sapien
39	231	14.8	248	1 NKPI_RAT	P18291 rattus norv
40	228.5	14.6	247	1 GRV6_MOUSE	P04187 mus musculu
41	221.5	14.2	254	1 TRV6_HUMAN	Q90903 homo sapien
42	221	14.2	261	1 TRV6_MOUSE	P07056 mus musculu
43	220	14.1	248	1 TRV6_MOUSE	Q90877 homo sapien
44	219	14.0	246	1 MCT4_RAT	P97592 rattus norv
45	218	14.0	258	1 VSP2_AGRAC	Q918x1 agkistrodon

## ALIGNMENTS

### RESULT 1

TRV2\_CHICK ID TRV2\_CHICK STANDARD; PRT: 248 AA.

AC Q90628; DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Trypsin I-P38 precursor (EC 3.4.21.4).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxId=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreeas;

RX MEDLINE=95251611; PubMed=7733885;

RA Wang K., Gan L., Lee I., Hood L.E.;

RT "Isolation and characterization of the chicken trypsinogen gene

family."

RL Biochem. J. 307:471-479 (1995).

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE

LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
EMBL: U15156; AAA79913.1; -  
PIR: S55067; S55067.  
HSSP: P00763; IDPO.  
DR MEROPS: S01.258; -  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin\_1.  
DR SMART: SM00020; Tryp\_Spc; 1.

[illegible]

Oy	660	GTTGGGACCTTCATGGAGAGGAGCGTGGCATCACCAATGTTCATAATATGCC	719
Dd	220	TTeelyCySaLaGlnmYsgIlyTrProGlyValThrylThrySValCySaenITyValSer	239
Oy	720	TGGATTGAGAACA	734
Dd	240	TripletysThrThr	244
RESULT 2			
RY1-CHICK			
ID	TRY1-CHICK	STANDARD;	PRT; 248 AA.
AC	Q90627;		
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DS	Trypsin I-P1 precursor (EC 3.4.21.4).		
De	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Pancreas;		
RX	MEDLINE=95251611; PubMed=7733885;		
RA	Wang K., Gan L., Lee I., Hood L.E.;		
RT	"Isolation and characterization of the chicken trypsinogen gene family.";		
RL	Biochem. J. 307:471-479(1995).		
CC	-1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-Xaa, Lys-Xaa.		
CC	-1- SUBCELLULAR LOCATION: Extracellular.		
CC	-1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcement/ or send an email to license@ebi.ac.uk).		
CC	-----		
DR	EMBL; U15155; AAA79912.1; .-		
DR	PIR; S55067; S55067.		
DR	HSSP; P00763; IDPO.		
DR	MEROPS; S01.151; .-		
DR	InterPro; IPRO01314; Chymotrypsin.		
DR	InterPro; IPRO01254; Ser_protease_Try.		
DR	Pfam; PF00089; Trypsin_1.		
DR	PRINTS; PRO0722; CHYMOTRYPSIN.		
DR	SMART; SMO020; TRY_SPC; 1.		
DR	PROSITE; PS50240; TRYPsin DOM; 1.		
DR	PROSITE; PS00134; TRYPsin_HIS; 1.		
DR	PROSITE; PS00135; TRYPsin_SER; 1.		
KM	Hydroxylase; Serine protease; Digestion; Pancreas; Zymogen; Signal; Multigene family.		
FT	SIGNAL	1	15
FT	PROPEP	16	25
FT	CHAIN	26	248
FT	ACT_SITE	65	65
FT	ACT_SITE	109	109
FT	ACT_SITE	202	202
FT	DISULFID	32	162
FT	DISULFID	50	66
FT	DISULFID	134	235
FT	DISULFID	141	208
FT	DISULFID	173	187
FT	DISULFID	198	222
FT	SITE	196	196
SEQ	SEQUENCE	248 AA;	26069 MM; CACFSB9912B23D99 CXC64;



## Alignment Scores:

Pred. No.: 2,05e-26 Length: 248  
 Score: 375.50 Matches: 87  
 Percent Similarity: 53.06% Conservative: 43  
 Best Local Similarity: 35.51% Mismatches: 100  
 Query Match: 24.07% Indels: 15  
 Gaps: 4

US-10-037-270-482 (1-866) x TRY1\_CHICK (1-248)

```

QY 42 ATGAAATAT-----GTCCTTATTTGGGCTGCTCGTGGACATTTTCTTGTCT 92
Db 1 Metlyserleuvalleuvalalaphavalgllyalthrval--Alapheproilser 19
QY 93 GACATCATCTGTTCGAAA-----GAAAGCCCTGCTCCCTAT 128
Db 20 AspGluabapbapblyllevallgllytysercysalaarvserralaalaprotyr 39
QY 129 TTGGTGTACCTCAAGTCTCACTTCAACCCCTGTGGGCGTCTCATCAAAACCGAGCTGG 188
Db 40 GlNvalserleuanserlglytyrhisrhepyselglyserleuilleserelintip 59
QY 189 GTGCTGCCCCAGCTCACTGCTATTACCAAAATGAAAGTGAATGCTGGAAATTTCAG 248
Db 60 Valleuseralaalahlscystyllyserserileginallybleglyglutyrasn 79
QY 249 AGCAGATCAGAGACGCTACTGACACACAAATTAAACCCATTCAGATCGCTGCTACTGG 308
Db 80 Leuallalaglnabpysergluthrleiserseerlyvalillearghiser 99
QY 309 AACTACATCATATCGCCGCCACAGATGACCTCACTGCTCACTGCTGCTGCTGCTGCC 368
Db 100 GLytyrasmalalaenthrleuanshspillemetleuileylsleerlyalala 119
QY 369 ATGCTCAATCCCAAGTCCAGCCCTTCCTCCCTGCCACCAATGTCAGGCGAGCACT 428
Db 120 Thrleuanserlyrvalasnthrvalproleuprothrsercysvalthralgllythr 139
QY 429 GTCTGTACTCTTCAAGTTTGAGTGGAGCCAGCAAAACAGTGGCCAGCTGACTTG 488
Db 140 Thrcysleuilleserlglytrpglyasnthrleuserseerlyserlytrproaspval 159
QY 489 CGGCAAGACCTGAGAGCCCGCCGATGCTGATGAGAAATGCCAAAAAGAACAGAAAGGA 548
Db 160 LeuGlnCysleuansalaprovalleuserseerlglnyserserlalytrpogly 179
QY 549 AAAAGCCACAGAAATTCCTTATGTGTGAATTTGTGAAGATTCAGCCGAATTTTGGG 608
Db 180 Arg1lethrseranmetilecysilleglytyrleuansnglylylvabpsercysgln 199
QY 609 GAGGTGCGCGTGTCTACTGCTCATTCGAAAGACAGCTCCAGGGAATC-----GAG 659
Db 200 G1yabpserlglytyrprovalvalcysabnnglylnleuglnglyllevalsertrp 219
QY 660 GTGGGACATTCATGAGGAGGAGGAGTGGGATTCACCAATGTTTCAATATGATTC 719
Db 220 lileglycysalaglnlysglytyrproglyalvalthrlyvalcysaentyrvalser 239
QY 720 TGGATTGAGAACT 734
Db 240 Tripleleysthrthr 244

```

## RESULT 3

TRY3 HUMAN STANDARD; PRT; 304 AA.  
 ID P35030; Q15665; OSUOVA;  
 AC 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Trypsin III precursor (EC 3.4.21.4) (Brain trypsinogen)  
 DE (Mecrotrypsinogen) (Trypsin IV).  
 GN PRSS3 OR TRY3 OR PRSS4 OR TRY4.  
 OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE=Brain;
RA MEDLINE=94123994; PubMed=8294000;
RX Wiegand U., Corbach S., Mann A., Kang J., Mueller-Hill B.;
RT "Cloning of the cDNA encoding human brain trypsinogen and
RT characterization of its product."
RL Gene 136:167-175(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM C).
RC TISSUE=Pancreas;
RX MEDLINE=90221895; PubMed=2326201;
RA Tani T., Kawashima I., Mita K., Takiguchi Y.;
RT "Nucleotide sequence of the human pancreatic trypsinogen III cDNA."
RL Nucleic Acids Res. 18:1631-1631(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM C).
RA Pukucka S.;
RL Submitted (FE8-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) (ISOFORM A).
RX MEDLINE=21686310; PubMed=11827488;
RA Katona G., Berglund G.I., Hajdu J., Graf L., Szilagyi L.;
RT "Crystal structure reveals basis for the inhibitor resistance of human
RT brain trypsin."
RL J. Mol. Biol. 315:1209-1218(2002).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-Xaa, Lys-Xaa.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=A;
CC IsoId=P35030-1; Sequence=Displayed;
CC Name=B;
CC IsoId=P35030-2; Sequence=VSP_005409;
CC Name=C;
CC IsoId=P35030-3; Sequence=VSP_005410.
CC -1- TISSUE SPECIFICITY: Pancreas and brain.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: X72781; CAB58178.1; -
CC EMBL: X71345; CA50484.1; -
CC EMBL: X15505; CA13527.1; -
CC EMBL: D45417; BAA08257.1; -
CC PIR: S33496; S33496.
CC PDB: 1H4W; 15-FEB-02.
CC MEROPS: S01.174; -.
CC GeneW: HGNC:9486; PRSS3.
CC GO: GO:0005576; C:extracellular; NAS.
CC GO: GO:0004295; F:trypsin activity; NAS.
CC GO: GO:0006508; P:proteolysis and peptidolysis; NAS.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Ser_protease_Try.
CC Pfam: PF00089; Trypsin_1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; TRYP_SPC; 1.
CC PROSITE: PS00240; TRYP_SIN_DOM; 1.
CC PROSITE: PS00134; TRYP_SIN_HIS; 1.
CC PROSITE: PS00135; TRYP_SIN_SER; FALSE_NEG.
CC HydroLase: Serine protease; Digestion; Zymogen; Signal;
CC Multigene family; Alternative splicing; 3D-structure.
CC SIGNAL 1 ?
CC PROPEP ? 80 ACTIVATION PEPTIDE.

```









QY 699 ATGTTTACAAATATGATCTGATTGAGAACCTGCTAAGAC 743  
 Db 231 LygValCysAenTyRValAspTrpIleGlnAsnThrIleAlaAsp 245

## RESULT 6

TRY1\_CANFA STANDARD; PRT; 246 AA.

DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin, cationic precursor (EC 3.4.21.4).  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 NC NCBI\_TaxID=9615;

RL [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=66284628; Pubmed=3841794;  
 RA Pinsky S.D., Laforge K.S., Scheele G.;  
 RT "Differential regulation of trypsinogen mRNA translation: full-length  
 RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes  
 RT in the dog pancreas";  
 RL Mol. Cell. Biol. 5:2669-2676(1985).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-Xaa, Lys-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M11590; AAA30900.1; -.  
 DR PIR; B26723; TRDGC.  
 DR HSSP; P00761; 1EPT.  
 DR MEROPS; S01.151; -.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR SMART; SM00020; TRYP\_SPC; 1.  
 DR PROSITE; PSS0240; TRYP\_SIN\_DOM; 1.  
 DR PROSITE; PSS0134; TRYP\_SIN\_HIS; 1.  
 DR PROSITE; PSS00135; TRYP\_SIN\_SER; 1.  
 KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KW Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23 ACTIVATION PEPTIDE.  
 FT CHAIN 24 246 TRYPSIN, CATIONIC.  
 FT ACT\_SITE 63 63 CHANGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 107 107 CHANGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 200 200 CHANGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 30 160 BY SIMILARITY.  
 FT DISULFID 48 64 BY SIMILARITY.  
 FT DISULFID 132 233 BY SIMILARITY.  
 FT DISULFID 139 206 BY SIMILARITY.  
 FT DISULFID 171 185 BY SIMILARITY.  
 FT DISULFID 196 220 BY SIMILARITY.  
 FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 246 AA; 26170 MW; E9E5A1DE2391BBB CRC64;

## Alignment Scores:

Pred. No.: 3,94e-24 Length: 246  
 Score: 350.50 Matches: 81  
 Percent Similarity: 50.83% Conservative: 42  
 Best Local Similarity: 33.47% Mismatches: 108  
 Query Match: 22.47% Indels: 11  
 DB: 1 Gaps: 2

US-10-037-270-482 (1-866) x TRY1\_CANFA (1-246)

QY 42 ATGAAATATGATCTTATTTGGGTGCTGCTGGACATTTTCTTTGCTGATCATCT 101  
 Db 1 MetIleThrPheIlePheLeuAlaLeuGlyAlaThrValAlaPheProIleAsp 20  
 QY 102 GTTCAGAA-----GAAACCTTCTCTCTTATTTGGTGTAC 137  
 Db 21 AspAspLysIleValGlyIleYrThrCysSerArgAsnSerValProYrGlnValSer 40  
 QY 138 CTCAGTCTCATTCACCCCTGTGGGGCTGCTCATCAACCCAGCTGGTCTGGCC 197  
 Db 41 LeuAsnSerGlyYrHisPheCysGlyGlySerLeuIleAsnSerGlnTrpValValSer 60  
 QY 198 CCAGCTCATCTGCTATTATTCACCAATTCGAAAGTATGCTGGAAATTTCAAGACAGTCTC 257  
 Db 61 AlaAlaHisCysEtyrLysSerArgIleGlnValArgLeuGlyIleYrAsnIleAlaVal 80  
 QY 258 AGAAGCGGTACTGAACAGACATTTAAACCCATTACAGTCTCGCTACTGAACTACAGT 317  
 Db 81 SerGlnGlyGlyGlyGlnPheIleAsnAlaAlaYserIleIleArgHisProArgYrAsn 100  
 QY 318 CATAAGGCCCCACAGATGACCTTCATGCTCATGAGCTGCTCAAGCTGCTGATCAT 377  
 Db 101 AlaAsnThrIleAspAsnAspIleMetLeuIleLysLeuSerSerProAlaThrLeuAsn 120  
 QY 378 CCCAAGTCCAGCCCTTCCCTCGCCACCAATGTCAGGCCAGGCACTGTCTGTCTA 437  
 Db 121 SerArgValSerAlaIleAlaLeuProLysSerCysProAlaAlaGlyThrGlnCysLeu 140  
 QY 438 CTCTCAGTTTGTGACTGCGAAGCCAAAGAAACAGTGGCCGACCTTGCCTGCGAGAAC 497  
 Db 141 IleSerGlyTrpIleYrAsnThrGlnSerIleGlyIleAsnTyRProAspValLeuGlnCys 160  
 QY 498 CTGAGAGCCCTGATGCTGTATTCAGAGATGCCAAAAACAGAACAGAAAAAGCCAC 557  
 Db 161 LeuIleAlaProIleLeuSerAspSerValCysArgAsnAlaTyRProGlyGlnIleSer 180  
 QY 558 AGGAATTCCTTATGCTGGAATTTGTAATGTAAGCCGAATTTTGGGGAGTGGCC 617  
 Db 181 SerAsnMetCysLeuGlyYrMetGlnGlyIleYrAspSerCysGlnGlyAspSer 200  
 QY 618 GTTGCTACTGTCATCTGCAAGAACAGCTCCAGGGAATC-----GAGTGGGGCAC 668  
 Db 201 GlyIleProValValCysAsnGlyIleGlnGlyValValSerTrpGlyAlaGlyCys 220  
 QY 669 TTATATGAGAGGAGACGCTGCATCTACACCAATGTTCAAAATATGATCTGATTTGAG 728  
 Db 221 AlaGlnLysGlyLysProGlyValSerProLysValCysLysTyRValSerTrpIleGln 240  
 QY 729 AACACT 734  
 Db 241 GlnThr 242  
 RESULT 7  
 TRY1\_XENLA STANDARD; PRT; 243 AA.  
 ID TRY1\_XENLA  
 AC P19759;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin precursor (EC 3.4.21.4).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 NC NCBI\_TaxID=8355;  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Pancreas;  
 RX MEDLINE=91007255; Pubmed=2210372;  
 RA Shi Y.B., Brown D.D.;  
 RT "Developmental and thyroid hormone-dependent regulation of pancreatic  
 RT genes in Xenopus laevis";  
 RT Genes Dev. 4:1107-1113(1990).

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL: X53458; CA37538.1; -.  
 DR PIR: A35871; A35871.  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.151; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PRO0722; CHYMOTRYPSIN.  
 DR SMART: SM00090; TRYP\_SPC; 1.  
 DR PROSITE: PS50240; TRYP\_SIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYP\_SIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYP\_SIN\_SER; 1.  
 KW Hydroxylase; Serine protease; Digestion; Zymogen; Signal;  
 KW Multigene family.  
 FT SIGNAL 1 15 BY SIMILARITY.  
 FT PROPEP 16 20 ACTIVATION PEPTIDE.  
 FT CHAIN 21 243 TRYPSIN.  
 FT ACT\_SITE 60 60 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 104 104 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 27 157 BY SIMILARITY.  
 FT DISULFID 45 61 BY SIMILARITY.  
 FT DISULFID 129 230 BY SIMILARITY.  
 FT DISULFID 136 203 BY SIMILARITY.  
 FT DISULFID 168 182 BY SIMILARITY.  
 FT DISULFID 193 217 BY SIMILARITY.  
 FT SITE 191 191 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 243 AA; 25492 MW; C5B8345A8B3F8031 CRC64;

Alignment Scores:  
 Pred. No.: 5.39e-24 Length: 243  
 Score: 349.00 Matches: 78  
 Percent Similarity: 51.88% Conservative: 46  
 Best Local Similarity: 32.64% Mismatches: 107  
 Query Match: 22.37% Indels: 8  
 DB: 1 Gaps: 2

US-10-037-270-482 (1-866) x TRY1\_XENLA (1-243)

QY 42 ATGAATAATATCTCTTCTTTGGGCTGCTGCGGACATTTTCTTGGCTCATCT 101  
 DB 1 MetlypheuleuLeuCyValleuLeuLylalaalalpheaspAspLys 20  
 QY 102 GTT-----CAGAAAGAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 146  
 DB 21 IlelleelglYalaltrCyValalysSerValProlyrilevalSerleuanser 40  
 QY 147 CACTTCAACCCCTGTGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 206  
 DB 41 GlylyrhisphecySglYglYserleuIlethraengIntrValalserAlaIahis 60  
 QY 207 TGCATTATTCACCAATCTGAAGATGATGCTGGGAATTTCAAGACAGACAGTACAGAGCT 266  
 DB 61 CyetYrlysaIaserIleGlnValaIrgleuLyluhIasenuIlealaleuSerGluGly 80  
 QY 267 ACTGAACAGCAATTAACCCCATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 326  
 DB 81 ThrIugInpheIleSerSerSerlyValIlealrghIserGlyTyraSerTyThr 100  
 QY 327 CCACGAGATGACCTCATGCTCATCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 386  
 DB 327 CCACGAGATGACCTCATGCTCATCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 386

DB 101 LeuaspAsnaspIleMetleuIleLysLeuSerSerProAlaSerleuAnaIaIaVal 120  
 QY 387 CAGCCCTTCCTCCCTGGCAGCACCAATGTCAGGCGACACTGCTGTCTACTCTGAGT 446  
 DB 121 AsnThrValProleuProserGlyCySerAlaIaGlyThrserysleuIleSerGly 140  
 QY 447 TTGAGCTGAGCAAGAAAACAGTGGCCGACACCTGACTTGGCGAGAACCTGAGGCC 506  
 DB 141 TrpGlyAsnThrleuSerenglySerAsnTyProAspIleuengIncyLeuAnaIa 160  
 QY 507 CCCGTGATGCTCATGAGAAATCCCAAAAACGAACAGAAAGAAAACCCAGAGATTCC 566  
 DB 161 ProIleuThrAsnIaGIncyAsnSerAlaTyProGlyIuIleThraIaAsnMet 180  
 QY 567 TTATGTGTGAATTTGTGAAAGATTCAGCCGAAATTTTGGGAGGTGGCGTGTCTACT 626  
 DB 181 IleCyValalGlyTrMetGluGlyLysAspSerCyGlnIuLyspserGlyGlyPro 200  
 QY 627 GTCATCTGCAAGACAAACAGCTCCAGGAGATC-----GAGGTGGCGCACTTCATGGA 677  
 DB 201 ValValCyasndlyGInuengInglyValalserTrpGlyTyrglyCySalameValg 220  
 QY 678 GGGAGCTGCGCATCTACACCAATGTTTACAAATATGATCTCGATTGAGAACACT 734  
 DB 221 AsnTyProGlyValTyThrLysValCyAsnTyAsnAlaIleIleGlnAsnThr 239

RESULT 8  
 TRY1\_RAT STANDARD; PRT; 246 AA.  
 ID TRY1\_RAT  
 AC P00762;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin I, anionic precursor (EC 3.4.21.4) (Pretrypsinogen I).  
 GN TRY1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=82265624; PubMed=6896710;  
 RA McDonald R.J., Stary S.J., Swift G.H.;  
 RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide  
 RT sequences of the cloned cDNAs".  
 RL J. Biol. Chem. 257:9724-9732(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85054880; PubMed=6094547;  
 RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,  
 RA Rutter W.J.;  
 RT "Structure of two related rat pancreatic trypsin genes".  
 RL J. Biol. Chem. 259:14255-14264(1984).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=87292123; PubMed=3112942;  
 RA Sprang S., Standing T., Pieterick R.J., Stroud R.M., Finer-Moore J.,  
 RA Xiong N.-H., Hamlin R., Rutter W.J., Craik C.S.;  
 RT "The three-dimensional structure of Asn102 mutant of trypsin: role of  
 RT Asp102 in serine protease catalysis".  
 RL Science 237:905-909(1987).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- MISCELLANEOUS: THIS SEQUENCE REPRESENTS THE PRECURSOR OF THE MAJOR  
 CC FORM OF TRYPSIN PRODUCED BY THE ADULT PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

DR EMBL; V01273; CAA24580.1; -.  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).  
 DR EMBL; J00778; AAA98518.1; -.  
 DR PIR; B26577; TRRT1.  
 DR PDB; 1TRM; 15-JUL-93.  
 DR PDB; 2TRM; 16-JUL-98.  
 DR PDB; 1BRA; 30-APR-94.  
 DR PDB; 1BRB; 31-JUL-94.  
 DR PDB; 1BRC; 31-MAY-94.  
 DR MEROPS; S01.094; -.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PSS0240; TRYPsin DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KM Multigene family; 3d-structure.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23 ACTIVATION PEPTIDE.  
 FT CHAIN 24 246 TRYPsin I, ANIONIC.  
 FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 107 107 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 200 200 CHARGE RELAY SYSTEM.  
 FT DISULFID 30 160  
 FT DISULFID 48 64  
 FT DISULFID 132 233  
 FT DISULFID 139 206  
 FT DISULFID 171 185  
 FT DISULFID 196 220  
 FT SITE 194 194  
 FT STRAND 25 25  
 FT STRAND 28 29  
 FT TURN 32 33  
 FT TURN 36 37  
 FT STRAND 38 42  
 FT STRAND 46 54  
 FT TURN 55 56  
 FT STRAND 57 60  
 FT HELIX 62 64  
 FT STRAND 70 73  
 FT STRAND 77 77  
 FT TURN 78 79  
 FT STRAND 86 95  
 FT TURN 97 98  
 FT STRAND 100 100  
 FT TURN 101 104  
 FT STRAND 105 105  
 FT TURN 106 106  
 FT STRAND 109 113  
 FT STRAND 127 127  
 FT TURN 135 136  
 FT STRAND 138 143  
 FT STRAND 157 157  
 FT STRAND 159 165  
 FT HELIX 168 174  
 FT TURN 176 178  
 FT TURN 181 182  
 FT STRAND 183 186  
 FT TURN 189 190  
 FT STRAND 194 194  
 FT TURN 197 198  
 FT TURN 200 201  
 FT STRAND 203 206  
 FT TURN 207 208  
 FT STRAND 209 216  
 FT TURN 223 224  
 FT STRAND 227 231  
 FT HELIX 232 234  
 FT HELIX 236 244  
 FT TURN 245 246

REQUIRED FOR SPECIFICITY.

SO SEQUENCE 246 AA; 25959 MW; 6AFA0DAD11943FB5 CRC64;

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
9,14e-24	346.50	54.50%	36.02%	22.21%	1
Matches:	76	Conservative:	39	Mismatches:	3
				Indels:	1
				Gaps:	1

US-10-037-270-482 (1-866) x TRY1\_RAT (1-246)

QY 111 GAAGACCTGCTCCATTTGGTGTACTCAAGCTCACTCAACCTGTGTGGGCTC 170  
 DB 32 GlnHisSerValProIyrglnValSerIleuAenSerGlyTyrHisSerGlyGlySer 51  
 QY 171 CTCATCAACCCAGCTGGTGTGGGCCCACTCACTGTATTACCAATCGAAAGTG 230  
 DB 52 LeuIleAenAerGlnIrrValValSerAlaIleHisCysTyrIysSerArgIleGlnVal 71  
 QY 231 ATGCTGGGAATTTCAAGACAGAGTCAGAGAGCTAGTCAAGACAAATTAACCCATT 230  
 DB 72 ArgLeuGlyGlnHisAenIleAenValLeuGlnGlyAerGlnIleHisAenAlaIle 91  
 QY 291 CAGATGCTCCGCTACTGGAATCACTCACTAGCGCCCAAGAGATGACCTCATGCTCATC 350  
 DB 92 LysIleIleIleHisProAenTyrSerSerTyrThrIleuAenAenAerIleIle 111  
 QY 351 AAGCTGGCTAAGCTCCATGCTCAATCCCAAGTCCAGCCCTTCCCTGGCCACC 410  
 DB 112 LysLeuSerSerProValIysLeuAenAlaArgValAlaProValAlaLeuProSerAla 131  
 QY 411 AATGTACGCGCAGGACCTGTCTGTACTCTCAAGTTTGACCTGACGACCAAGAAACAT 470  
 DB 132 CysAlaProAlaGlyThrGlnCysLeuIleSerGlyTyrGlyAenThrLeuSerAenGly 151  
 QY 471 GGCGACACCTGACTGCGGCGAGAACCTGAGAGCCCGCTGATGTCTGATCGAATGC 530  
 DB 152 ValAenAenProAerProAerLeuGlnCysValAerAlaProValLeuSerGlnAlaAerCys 171  
 QY 531 CAAAAACAGAACAGAAAGAACAGCAAGCAATTCCTTATGTGTGAATTTGTGAAGTA 590  
 DB 172 GluAlaIleAlaTyrProGlyGlnIleThrSerSerMetIleCysValGlyPheLeuGlnGly 191  
 QY 591 TTCAGCCGAATTTTGGGAGGAGTGGCCGTTCCTACTGTATCTCGCAAGCAAGCTTCAG 650  
 DB 192 GlyIysAerSerCysGlnGlyAerSerGlyGlyProValValAerGlnGlnLeuGln 211  
 QY 651 GGAATC-----GAGGTGGGGCACTTCATGCGAGGAGCGGACGTCATCAACCAAT 701  
 DB 212 GlyIleValSerTyrGlyTyrGlyCysAlaLeuProAerAenProGlyValIyrrHisLys 231  
 QY 702 GTTTACAAATATGATCTGATCTGATGAAACACT 734  
 DB 232 ValCysAenPheValGlyTyrIleGlnAerThr 242  
 RESULT 9  
 TRY2\_CANFA  
 ID TRY2\_CANFA STANDARD; PRT; 247 AA.  
 AC P06872;  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin, anionic precursor (EC 3.4.21.4).  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 NC NCBI\_TaxID=9615;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=66284628; PubMed=3841794;  
 RA Pinsky S.D., Laforge K.S., Scheele G.,  
 RT "Differential regulation of trypsinogen mRNA translation: full-length

RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes  
 RT in the dog pancreas.";  
 RL Mol. Cell. Biol. 5:2669-2676(1985).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL/ M1589; AAA30899.1; -.  
 DR PIR: A26273; TRDC.  
 DR HSPSP; P00763; IDPO.  
 DR MEROPS; S01.258; -.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydroxylase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KW Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23  
 FT CHAIN 24 247  
 FT ACT\_SITE 63 63  
 FT ACT\_SITE 107 107  
 FT ACT\_SITE 200 200  
 FT DISULFID 30 160  
 FT DISULFID 48 64  
 FT DISULFID 132 233  
 FT DISULFID 139 206  
 FT DISULFID 171 185  
 FT DISULFID 196 220  
 FT SITE 194 194  
 SQ SEQUENCE 247 AA; 26423 MW; 374B9D31D6DB8EAF CMC64;  
 Alignment Scores:  
 Pred. No.: 1.02e-23 Length: 247  
 Score: 346.00 Matches: 76  
 Percent Similarity: 52.56% Conservative: 47  
 Best Local Similarity: 32.48% Mismatches: 103  
 Query Match: 32.18% Indels: 2  
 Gaps: 8  
 US-10-037-270-482 (1-866) x TRY2\_CANFA (1-247)  
 QY 57 TATTGGGTGTCCTCGGAGACATTTTCTTGTCTGATCATCTGTTACG----- 107  
 Db 9 PheLeuGlyAlaAlaValAlaThrProThraPaspAspLysIleValGlyTyr 28  
 QY 108 -----AAAGAAGACCTGCTCCCTATTGGTAGCACTCAAGCTCACTTAACCCCTG 161  
 Db 29 ThrCysGluLysSerValProTyrGlnValSerLeuAsnIleGlyTyrHisPheCys 48  
 QY 162 GTGGAGCTCTCATCAACACAGCTGGTGGTGGCCAGCTCACTGTTATTCAAAAT 221  
 Db 49 GlyGlySerLeuIleSerAspGlnTyrAlaValSerAlaIleHisCysTyrLysSerArg 68  
 QY 222 CTGAAGTGTATCTGGGAATTTCAAGACAGATCAGACACGGTACTGAACAGCAATT 281  
 Db 69 IleGlnValArgLeuGlyGlnTyrAsnIleAspValLeuGluGlnGlnPheIle 88  
 QY 282 AACCCATTAGATCGCGCTACGAGAACTACGAGTACGATACGCCGCCACAGAGTACCTC 341  
 Db 89 AsnSerAlaLysValIleArgHisIleProGlnTyrAsnSerTyrPheLeuAspAsnIle 108  
 QY 342 ATGCTCATCAAGCTGGCTAAGCTGCTCATGCTCATCCCAAGTCCAGCCCTTCCCTCC 401

Db 109 MetLeuIleLysLeuSerSerProAlaValLeuAsnAlaArgValAlaThrIleSerLeu 128  
 QY 402 GCCACACCAATGTACAGGACGAGCTGTGCTACTCTCAGTGTGACAGCCAA 461  
 Db 129 ProArgAlaCysAlaAlaPheGlyThrGlnCysLeuIleSerIleTyrPheAsnThrLeu 148  
 QY 462 AAAACAGTGGCCGACACCTGACTTGGCGAGACCTGAGGCCCCCGTATGTCTGAT 521  
 Db 149 SerSerGlyThrAsnTyrProGlnLeuGlnCysLeuAspAlaProIleLeuThrGln 168  
 QY 522 CGAATGCGCAAAAACAGAACAGAAAAAGCCAGAAATTCCTATGTGGAATTT 581  
 Db 169 ALGlnCysGlnAlaSerTyrProGlyGlnIleThrGlnAsnMetIleCysAlaGlyPhe 188  
 QY 582 GTGAAGTATTACGCCGAAATTTTGGGAGGTGGCCGCTGCTACTGTCATCTGCAAAAGAC 641  
 Db 189 LeuGlnGlyLysAlaPheSerCysGlnGlyAspSerCylGlyProValAlaCysAsnGly 208  
 QY 642 AAGCTCCAGGAATC-----GAGGTGGGCACTTCATGAGAGGAGCGTGGCATC 692  
 Db 209 GluLeuGlnGlyIleValSerTyrGlyTyrGlyCysAlaGlnLysAsnLysProGlyVal 228  
 QY 693 TACACCAATTTTACAAATATGTATCTGATGAGAACACT 734  
 Db 229 TyrThrLysValCysAsnPheValAspTyrPheIleGlnSerThr 242  
 RESULT 10  
 TRY3\_CHICK  
 ID TRY3\_CHICK STANDARD; PRT; 248 AA.  
 AC Q90629;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin II-P29 precursor (EC 3.4.21.4).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=95251611; PubMed=7733885;  
 RA Wang K., Gan L., Lee I., Hood L.E.;  
 RT "Isolation and characterization of the chicken trypsinogen gene  
 RL family." J. 307:471-479(1995).  
 RL Biochem. J. 307:471-479(1995).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE  
 CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL/ U15157; AAA79914.1; -.  
 DR PIR: S55066; S55066.  
 DR HSPSP; P00763; IDPO.  
 DR MEROPS; S01.151; -.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydroxylase; Serine protease; Digestion; Pancreas; Zymogen; Signal;





DR PDB; 1F5R; 04-JUL-01.  
DR PDB; 1F7Z; 04-JUL-01.  
DR PDB; 1FV8; 04-JUL-01.  
DR PDB; 1J14; 11-FEB-03.  
DR PDB; 1J15; 11-FEB-03.  
DR PDB; 1J16; 11-FEB-03.  
DR PDB; 1J17; 11-FEB-03.  
DR PDB; 1QL9; 25-AUG-00.  
DR PDB; 1TRM; 15-JUL-93.  
DR PDB; 2TRM; 16-JUL-88.  
DR PDB; 3TCK; 04-JUL-01.  
DR MEROS; S01.258; -.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; Trypsin; 1.  
DR SMART; SMO0020; Tryp\_Spc; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR HydroLase; Serine protease; Digestion; Pancreas; Zymogen; signal;  
KW Multigene family; 3D-structure.  
FT SIGNAL 1 15  
FT PROPEP 16 23 ACTIVATION PEPTIDE.  
FT CHAIN 24 246 TRYPSIN II, ANIONIC.  
FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 107 107 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 200 200 CHARGE RELAY SYSTEM.  
FT DISULFID 30 160  
FT DISULFID 48 64  
FT DISULFID 132 233  
FT DISULFID 139 206  
FT DISULFID 171 185  
FT DISULFID 196 220  
FT SITE 194 194  
FT CONFLICT 84 84  
FT CONFLICT 88 88  
FT STRAND 25 25  
FT STRAND 28 29  
FT TURN 32 33  
FT TURN 36 37  
FT TURN 37 38  
FT STRAND 46 42  
FT STRAND 48 54  
FT TURN 55 56  
FT STRAND 57 60  
FT HELIX 62 64  
FT STRAND 70 73  
FT STRAND 77 77  
FT TURN 78 79  
FT STRAND 86 95  
FT TURN 97 98  
FT TURN 101 103  
FT TURN 105 106  
FT STRAND 109 113  
FT STRAND 120 120  
FT TURN 121 122  
FT STRAND 123 123  
FT STRAND 127 127  
FT TURN 135 136  
FT STRAND 138 143  
FT STRAND 157 157  
FT STRAND 159 164  
FT HELIX 168 174  
FT TURN 176 178  
FT TURN 181 182  
FT STRAND 183 185  
FT STRAND 194 194  
FT TURN 197 198  
FT TURN 200 201  
FT STRAND 203 206  
FT TURN 207 208  
FT STRAND 209 216  
FT TURN 223 224  
FT STRAND 228 231  
FT HELIX 232 234

REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
N -> D (IN REF. 1).  
V -> I (IN REF. 1).

FT HELIX 236 246  
SQ SEQUENCE 246 AA; 26228 MW; A8D3630809AE606 CRC64;  
Alignment Scores:  
Pred. No.: 3.23e-23  
Score: 340.50  
Percent Similarity: 50.83%  
Best Local Similarity: 31.82%  
Query Match: 21.83%  
DB: 1  
Gaps: 2  
US-10-037-270-482 (1-866) x TRY2\_RAT (1-246)  
QY 42 ATGAATATCTCTTCAATTTGGTGGTCTGCGGACATTTTCTTGCTGCTCATCT 101  
Db 1 MetArgAlaLeuLeuPheLeuAlaLeuValGlyAlaAlaValAlaPheProValAsp 20  
QY 102 GTTCAGAA-----GAGACCTCTCTCTTATTTGGTTAC 137  
Db 21 AspAspLysIleValGlyGlyTyrThrCysGlnGlnLysSerValProTyrGlnValSer 40  
QY 138 CTGAGCTCACTTCAACCCCTGTGGCGCTCTCATCAACCCAGCTGGTGGTGGCC 197  
Db 41 LeuAsnSerGlyTyrHisPheCysGlyGlySerLeuIleAsnAspGlnTyrValSer 60  
QY 198 CCACTCACTGCTATTTCACCAATCTGAAGATGATGCTGGAATTTCAAGACAGATC 257  
Db 61 AlaAlaHisCysTyrTyrSerArgIleGlnValArgLeuGlyGlnHisAsnIleAsnVal 80  
QY 258 AGAGACGTACTGAACAGACAAATTAACCCATTCAAGTCTCGCTCACTGAACACTAGT 317  
Db 81 LeuGlnGlyAsnGlnGlnPheValAsnAlaAlaIleIleLeuHisPheAsnAsp 100  
QY 318 CATAGCGCCCAAGATGACCTCATGCTCAAGTGGCTCAAGCTGCCATGCTCAAT 377  
Db 101 ArgLysThrLeuAsnAsnAspIleMetLeuIleLysLeuSerSerProValLysLeuAsn 120  
QY 378 CCCAAGTCCAGCCCTTCCCTGCGCCACCAATGTCAAGCCAGGCACTGCTGTCTA 437  
Db 121 AlaArgValAlaThrValAlaLeuProSerSerCysAlaProAlaGlyThrGlnCysLeu 140  
QY 438 CTCTCAGTTTGGACTGGAGCCCAAGAAACAGTGGCCGACACCTTGACTGGCGAGAAC 497  
Db 141 IleSerGlyTyrGlyAsnThrLeuSerSerGlyValAsnGlnProAspLeuGlnCys 160  
QY 498 CTGAGGCGCCCTGATGCTGTGATCGAATGCCAAATAACAGAAACAAGAAAGCCAC 557  
Db 161 LeuAspAlaProLeuLeuProGlnAlaAspCysGlnLysSerTyrProGlyLysIleThr 180  
QY 558 AGGAATTCCTTATGTGGAATTTGTGAAGTATTCAGCCGAATTTTGGGAGGTGGCC 617  
Db 181 AspAsnMetValCysValGlyPheLeuGlnGlyGlyLysAspSerCysGlnGlyAspSer 200  
QY 618 GTTGCTACTGTCACTGCAAGACCAAGCTCAGGGGAATC-----GAGTGGGGCAC 668  
Db 201 GlyGlyProValAlaCysAsnGlyGlnLeuGlnGlyIleValSerTyrGlyTyrCys 220  
QY 669 TTGATGGAGAGGAGCGTGGCATCTACCAATGTTTCAATATGATCTGATTTGAG 728  
Db 221 AlaLeuProAspAsnProGlyValTyrThrLysValCysAsnTyrValAspTyrIleGln 240  
QY 729 AACACT 734  
Db 241 AspThr 242  
RESULT 12  
ID TRYP\_PIG STANDARD; PRT; 231 AA.  
AC P00761;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Trypsin precursor (EC 3.4.21.4).

OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9623;  
 RN  
 RA SEQUENCE OF 1-10.  
 RA Charles M., Rovey M., Guidoni A.A., Desnuelle P.;  
 RT "on trypsinogen and trypsin of pig.";  
 RL Biochim. Biophys. Acta 69:115-129(1963).  
 RN  
 RP  
 RP SEQUENCE OF 9-231.  
 RA MEDLINE=73258692; Pubmed=4738933;  
 RA Herndon M.A., Ericsson L.H., Neurath H., Walsh K.A.;  
 RT "determination of the amino acid sequence of porcine trypsin by  
 RT sequenator analysis.";  
 RL Biochemistry 12:3146-3153(1973).  
 RN  
 RN  
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=93187998; Pubmed=8445634;  
 RA Huang Q., Liu S., Tang Y.;  
 RT "refined 1.6-A resolution crystal structure of the complex formed  
 RT between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the  
 RT equaen family. Detailed comparison with bovine beta-trypsin and its  
 RT complex.";  
 RL J. Mol. Biol. 229:1022-1030(1993).  
 RN  
 RN  
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=92201369; Pubmed=1551419;  
 RA Huang Q., Liu S., Tang Y., Zeng F., Qian R.;  
 RT "amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray  
 RT crystal structure of its complex with porcine beta-trypsin.";  
 RL FEBS Lett. 297:143-146(1992).  
 RN  
 RN  
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=95035057; Pubmed=7947985;  
 RA Huang Q., Wang Z., Li Y., Liu S., Tang Y.;  
 RT "Refined 1.8-A resolution crystal structure of the porcine epsilon-  
 RT trypsin.";  
 RL Biochim. Biophys. Acta 1209:77-82(1994).  
 RN  
 RN  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.  
 RX MEDLINE=97390427; Pubmed=9242660;  
 RA Stubbs M.T., Morenweiser R., Sturzebecher J., Bauer M., Bode W.,  
 RA Huber R., Piechotka G.P., Matschner G., Sommerhoff C.P., Fritz H.,  
 RA Auerwald E.A.;  
 RT "The three-dimensional structure of recombinant leech-derived  
 RT trypsin inhibitor in complex with trypsin. Implications for the  
 RT structure of human mast cell trypsin and its inhibition.";  
 RL J. Biol. Chem. 272:19931-19937(1997).  
 RN  
 RN  
 RN X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.  
 RX MEDLINE=98046095; Pubmed=9384562;  
 RA di Marco S., Priestle J.P.;  
 RT "Structure of the complex of leech-derived trypsin inhibitor (LDTI)  
 RT with trypsin and modeling of the LDTI-trypsin system.";  
 RL Structure 5:1465-1474(1997).  
 CC  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR  
 DR PDB; 1MCT; 31-JAN-94.  
 DR PDB; 1AKS; 12-FEB-97.  
 DR PDB; 1EPT; 07-FEB-95.  
 DR PDB; 1TFX; 21-JAN-98.  
 DR PDB; 1LDT; 20-MAY-98.  
 DR PDB; 1ANI; 01-JUL-98.  
 DR PDB; 1AVW; 18-NOV-98.  
 DR PDB; 1AVX; 18-NOV-98.  
 DR PDB; 1C9P; 26-SEP-01.  
 DR PDB; 1DF3; 30-NOV-99.  
 DR PDB; 1EJA; 02-MAR-01.  
 DR PDB; 1EWJ; 28-JUN-00.

DR PDB; 1FMG; 07-NOV-01.  
 DR PDB; 1FN6; 07-NOV-01.  
 DR PDB; 1FNT; 07-NOV-01.  
 DR PDB; 1LT2; 05-JUN-02.  
 DR PDB; 1QOU; 14-JUN-00.  
 DR MEROPS; S01.151;  
 DR InterPro; IPR001254; Ser protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; TRYR\_SPC; 1.  
 DR PROSITE; PSS0240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KM Hydroxylase; Serine protease; Digestion; Pancreas; Zymogen;  
 KM 3D-structure.  
 FT PROPEP 1 8  
 FT CHAIN 9 231  
 FT ACT\_SITE 48 48  
 FT ACT\_SITE 92 92  
 FT ACT\_SITE 185 185  
 FT DISULFID 15 145  
 FT DISULFID 33 48  
 FT DISULFID 117 218  
 FT DISULFID 124 191  
 FT DISULFID 156 170  
 FT DISULFID 181 205  
 FT SITE 179 179  
 FT VARIANT 20 20  
 FT STRAND 10 10  
 FT STRAND 13 14  
 FT TURN 17 18  
 FT TURN 21 22  
 FT TURN 23 27  
 FT STRAND 31 39  
 FT TURN 40 41  
 FT STRAND 42 45  
 FT STRAND 47 49  
 FT HELIX 55 58  
 FT STRAND 62 62  
 FT STRAND 63 64  
 FT TURN 71 80  
 FT TURN 82 83  
 FT TURN 86 88  
 FT TURN 90 91  
 FT STRAND 94 98  
 FT STRAND 112 112  
 FT TURN 120 121  
 FT STRAND 123 128  
 FT STRAND 142 142  
 FT STRAND 144 150  
 FT STRAND 153 159  
 FT HELIX 161 163  
 FT TURN 166 167  
 FT STRAND 168 171  
 FT TURN 174 175  
 FT STRAND 179 179  
 FT TURN 182 183  
 FT STRAND 185 186  
 FT TURN 192 193  
 FT STRAND 194 202  
 FT STRAND 207 207  
 FT TURN 208 209  
 FT STRAND 210 210  
 FT STRAND 212 216  
 FT HELIX 217 219  
 FT HELIX 221 231  
 SQ SEQUENCE 231 AA; 24409 MW; A0A125CF7FC138C2 CRC64;  
 Alignment Scores:  
 Pred. No.: 9.14e-23  
 Score: 335.50  
 Percent Similarity: 52.89%  
 Best Local Similarity: 34.67%  
 Length: 231  
 Matches: 78  
 Conservative: 41  
 Mismatches: 97

REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 I -> V.

```

Query Match: 21.51% Indels: 9
DB: 1 Gaps: 2

US-10-037-270-482 (1-866) x TRYP_PIG (1-231)

QY CTGCGTGGGACATTTTCTTTTGTGCTGACATCTGTTCAGAAAAGAACCCGTCCCTAT 128
   ::: ||| :::: ||| ||| :::: ||| ||| :::: ||| ||| :::: ||| |||
Db 9 lIeValGlYlYlYrThrCYsAlAlAaAnSerIle-----ProTyr 22

QY 129 TTGGTGTACTCAAGTCTCATCTTCAACCCCTGTGTGGCGTCTTCATCAACCAAGCTG 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 GlNValSerLeuAnSerGlySerIlePheCYsGlYlYSerLeuIleAnSerGlnTrp 42

QY 189 GTGCTGGCCCCAGCTACATCTTATTACCAATCTGMAAGTGAAGTCTGGAAATTTCAG 248
   ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||
   43 ValValSerIleAlaIleHicSYrYrLYsSerArgIleGlnValArgLeuGlYGlNHIAsn 62

QY 249 AGCAGAGTACGACGCGTACTGTACGACAGAAATTAACCCATTCCAGATGTCGCGTAC 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 lIeAspValleuGlYleuGlYAnGlnGlnPheIleAlaAlAlValSYrIleThHicPro 82

QY 309 AACTCAAGTATPAGCGCCCCACAGATGACCTCATGCTCATCAAGCTGGCTTAAGCTGCC 368
   ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||
Db 83 AsnIleAsnGlnYAnThrLeuAspAsnAspIleMetLeuIleLYsLeuSerSerProAla 102

QY 369 ATGCTCATCTCCAAAGTCCAGCCCTTCCCTTCGCCACCAACCATGTAGCGCCAGGACT 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 ThrLeuAsnSerArgValAlaThrValSerLeuProArgSerCYsAlAlAlaGlYThr 122

QY 429 GTCTGTACTACTCAGATTTTGGACCTGGACGACAAAGAACAGTGGCCGACCTGACTGT 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 GlUCysAluIleSerGlyTrpAlYAnThrLYsSerSerGlySerTyrProSerLeu 142

QY 489 CGGCGAAGACTGAGAGGCCCCCGCTGATGTGTGATCGAAGATGCAAAAAAGAAACAAAGA 548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 LeuGlnCYsLeuLYsAlaProValIleuSerAspSerSerCYsLYsSerSerTyrProGly 162

QY 549 AAAAGCCACAGGAATTCCTTAGTGTGGAATTTGTGAAAGTATTCAGCCGAATTTTGGG 608
   ::: ||| ||| :::: ||| ||| :::: ||| ||| :::: ||| ||| :::: |||
Db 163 GlNlIeThrLYsAnMetIleCYsValGlYPheLeuGlYGlYLYsAspSerCYsGln 182

QY 609 GAGGTGGCCGTTGCTCTACTGTCATCTGACCAAGAACAGCTCCAGGGGAATC-----GAG 659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 GlYAspSerIleGlyProValValCYsAnGlnGlnLeuGlnGlyIleValSerTrpGly 202

QY 660 GTGGGCGACCTTCATGGAGGCGGACGTGGCATCTACACCAATGTTTCAATATGTAATCC 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 TyrGlYCYsAlaGlNLYsAnLYsPProGlyValIYrThrLYsValCYsAnThYValAsn 222

QY 720 TGGATTGAGAACT 734
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 TrpIleGlnGlnThr 227

RESULT 13
RTRY2_HUMAN STANDARD; PRT; 247 AA.
AC P07478;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin II precursor (EC 3.4.21.4) (Anionic trypsinogen).
GN PRSS2 OR TRY2 OR TRYP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
PI [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86221712; PubMed=3011602;
RA Emt M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,
RA Matsubara K.
RT "Cloning, characterization and nucleotide sequences of two cDNAs
encoding human pancreatic trypsinogens.";
```

```

RL Gene 41.505-310(1986) .
[21]
RN SEQUENCE OF 16-49.
RX MEDLINE=90091010; PubMed=2598466;
RA Kmieland M., Ruessick C., Marks W.H., Borgstrom A.;
RT "Immunoreactive anionic and cationic trypsin in human serum.";
RL Clin. Chim. Acta 184:31-46(1989) .
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-Xaa, Lys-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement
CC or send an email to license@ebi.ac.uk).
-----
DR EMBL, M27602; AAA61232.1; -.
DR PIR, B25852; B25852.
DR HSPB, P00763; IDPO.
DR MEROPS, S01.258; -.
DR Genew, HGNC:9483; PRSS2.
DR MIM, 601564; -.
DR GO, GO:0004295; F:trypsin activity; TAS.
DR InterPro, IPR001254; Ser_protease_Try.
DR Pfam, PF00089; trypsin; 1.
DR SMART, SM00020; Tryp_Spc; 1.
DR PROSITE, PS00240; TRYPSIN_DOM; 1.
DR PROSITE, PS00134; TRYPSIN_HIS; 1.
DR PROSITE, PS00135; TRYPSIN_SER; 1.
DR K0M Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KM Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 23 ACTIVATION PEPTIDE.
FT CHAIN 24 247 TRYPSIN II.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26488 MW; 82B0F41EB8E3D5DB CRC64;

Alignment Scores:
Pred. No.: 1,14e-22 Length: 247
Score: 334.50 Matches: 75
Percent Similarity: 52.83% Conservative: 37
Best local Similarity: 35.38% Mismatches: 97
Query Match: 21.44% Indels: 3
Gaps: 1

US-10-037-270-482 (1-866) x TRY2_HUMAN (1-247)
QY 108 AAGAAGACCCCTGCTCCCTATTGGTGTACCTCAAGTCTCACTTCAACCCCTGTGGGC 167
Db 31 GUGUUAANSeValProGlyInValSerLeuAnSeRgLYrHISpHeCyGGLgLY 50
QY 168 GTCCCTAAACCCAGCTGGGTGCTGGCCCCAGCTCACTGTCATTTTAAACAAATGTGAAA 227
Db 51 SerLeuILSeRcJUGInInTrValValSerLacIgHIScYrLYrSeRArgILeGIn 70
QY 228 GGTATCTGTGGAATTTCAAGAGCAGAGTCAAGAGCGGTACTGAAAGACAAATTTACCCC 287
Db 71 ValArgLeuGlyGLuHISaSeNLeGILValLeuGluGluGluGluGluInPheILeAraLa 90
QY 288 ATTCAAGATGTCGGCTACTGGAATCAAGATAGATATAGACGCCCCACAGATGATCACTGATC 347
Db 91 AlAyLleILeILeArgHISrProLYrGlyrAnSeRArgHrHISLeuAraAnPILeLeuLeu 110

```

OY		348	ATCAAGCTGGGACCCATTGCAGTTCATGCCAAGAAGTACAACCCTTGCCGCAAC	407
Dd		111	IleIysLeuSerSeSProAlaValIleAsnSerArgValSeraIleSerLeuproThr	130
OY		408	ACCATGTTCAGCGCAGCACACTGTCTGTCTACTCTCAGAGTTTGGACTGGAACAA	467
Dd		131	ALAProProlalaIGLYThrcIUberLeuIIeSerGIlyTrpGLyAsnthrLeuSerseR	150
OY		468	AGTGCCCAACACCTTGACTTGGCGGAGAACCCTGGAGCCCCCGTAGATGTCTGATGAGAA	527
Dd		151	GlyalaaEPryPrOaspGlutenuInCySluaEapAlaPaProvallauSerGINaIglu	170
OY		528	TGCCAAAAAACGAACAAGAAAAGAACCAACAGACATACTTCTTAITGTCTGAATTTTGAAA	587
Dd		171	CysglualaseSTyr-ProgILyylserlthrnaasamebpheCYavaGIypHeLuglu	190
OY		588	GTATTACACCCGATTTTTGGGAGGCGCCGTGTGCTACTGTCATCTGCAAGACAAGCTC	647
Dd		191	GIglyLyvaSpSeRCyeGINclyaSpBerGIygLyprovalVaISerAmNcIguLeu	210
OY		648	CAGGGAATC-----GAGGTGGGCACTTCATGGAGGAGGACGTGGCATCTACAC	698
Dd		211	GIingilyllevalsetrpGLyTyrgLIcySaINaglnLyasaMdyProgILyValTythr	230
OY		699	AATGTTACAATAATGATATCTATCCGTGATTGAGAACAACACT	734
Dd		231	LysVALtyraNTyralaSprtIllelysAeprrH	242

RESULT 14  
TRY2\_XENLA STANDARD; PRT; 244 AA.

ID	_TRY2_XENLA	PRT;	244	AA.
AC	F70059,			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Trypsin precursor (EC 3.4.21.4)			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;			
OC	Xenopodinae; Xenopus.			
OK	NCHI_TaxID=8335;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Wang K., Lytle L., Gan L., Hood J.E.?			
RL	Submitted (SEP-1996) to the EMBL/Genebank/DBJ databases.			
CC	- CATALYTIC ACTIVITY: Preferential cleavage: Arg-Xaa, Lys-Xaa.			
CC	- SUBCELLULAR LOCATION: Extracellular			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration-			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isdb-sib.ch/announce/isdb-stb.ch).			
OR	or send an email to license@isb-sib.ch).			
DR	EMBL: U72330; AAB17274.1; --			
DR	HSSP: P00763; IDBO.			
DR	MEROFS_ S01.258; --			
DR	InterPro: IPR001314; CHymoTrypsin.			
DR	InterPro: IPRO01254; Ser_protease_Try.			
DR	pFam: PF00089; trypsin_1.			
DR	PRINTS: PR00722; CHYMOTRYPSIN.			
DR	SMART: SM00020; Tryp_SPC; 1..			
DR	PROSITE: PS0240; TRYP SIN_DOM; 1.			
DR	PROSITE: PS00134; TRYPS IN_HIS; 1.			
DR	PROSITE: PS00135; TRYPS IN_SER; 1.			
KW	Hydroxylase; Serine protease; Digestion; Zymogen; Signal; Multiogene family.			
SIGNAL	1			
FT PROPEP	16	15		BY SIMILARITY.
FT FT	21	21		ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN	22	244		TRYP SIN.

FT	ACT_SITE	61	61		CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	ACT_SITE	105	105		CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	ACT_SITE	198	198		CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	DISULFID	28	158		BY SIMILARITY.		
FT	DISULFID	46	62		BY SIMILARITY.		
FT	DISULFID	130	231		BY SIMILARITY.		
FT	DISULFID	137	204		BY SIMILARITY.		
FT	DISULFID	169	183		BY SIMILARITY.		
FT	DISULFID	194	218		BY SIMILARITY.		
FT	SITE	192	192		REQUIRED FOR SPECIFICITY (BY SIMILARITY).		
SEQ	SEQUENCE	244 AA;	26079 MW;	C63F29CB330B323 CRC64;			
Alignment Scores:							
Pred. No.:	2.14e-22	Length:	244				
Score:	331.50	Matches:	74				
Percent Similarity:	51.25%	Conservative:	49				
Best Local Similarity:	30.83%	Mismatches:	108				
Query Match:	21.25%	Indels:	9				
DB:	1	Gaps:	2				
US-10-037-270-482 (1-866) x TRY2_XENLA (1-244)							
OY	42	ATGAATAATGCTTCTATTGGTGCGTCGCCGTGGGACATTTTCTTGCTGCATCATC	101				
Db	1	MethylphenyleuvalIleleuvalIleuuglyAlalavalAlaphetulsparapR	20				
OY	102	-----GTTCAAGAAAGAACCCCTGCTCCCTATTGGTGCTACCTCAAG	143				
Db	21	LysIleValglgylPheThrCybaIalayaAmdAlaValProtyrGlInlaserleuaen	40				
OY	144	TCTCACTTCAACCCTGTGTGGGCTCTCATCAAAACCAGGTGGTGTGCCCCAGCT	203				
Db	41	AlaglyTrxHisphnCysglyGlyserIleuIleaenserglnPrvalValserAlala	60				
OY	204	CACGSGCTATTTACCAAACTGTAAAGTAGTGCGGGAAMTTTCAAGACAGAGTCAGAC	253				
Db	61	HiccsyryLysSerArgIleGlnAlaArgleuUlgLnHlaAmnIleAlaleuaenglu	80				
OY	264	GGAATCAAGACAGACATTAACCCCATTCAGATCGTCCGCTACGTGACATCACAGTACAGC	323				
Db	81	GlythrGluginPhelIleaspserrInyValIlelyvhiProahntyrAenseraAg	100				
OY	324	GCCCCACGATGACCTCATGCTCATCAAGCTGGCTAAAGCCTGCATGCTCAATCCCAA	383				
Db	101	AamLeuAparanaApRIleMetleuIleLysIeuSerThrThAlaArgIeuseralAan	120				
OY	384	GTCACGCCCTTCCCCTGCCACACCAATGTCACAGCCAGGACACTGTCTGTACTCTCA	443				
Db	121	IlegInservalProleuProserIaCybaIserralaglITThAsmCYseuIIeser	140				
OY	444	GATTGACGTGAGCAGAAGAAACAGTGGCCGACACCTGATCGGACAGAACTGGAG	503				
Db	141	GIlyTrpgLyAnthrInLysSerSerIlyThrrentyrProAspleuIeugInCybleuaen	160				
OY	504	GCCCCCGATGTCTGATCGAAGATGCCAAAAACAGAACAGAAAGAAAACCAAGAAAT	563				
Db	161	AlaProIleLeuthrAspserrGIncySerAensertyrProglYglunIleThrysaen	180				
OY	564	TCCTTATGTGGAATTTGTGAAGATATTACGCCCAATTTTTGGGAGGCTGGCCGTGCT	623				
Db	181	MetPhecYbaIaglyPheleuAlaglYglYusApsereCyseglncIyAspsertglYgl	200				
OY	624	ACTGCATCTGCAAGAAGAGCTCCAGGAATC-----GAGGTGGGACACTTCATC	674				
Db	201	ProvalValCybaenglyGlnleuInglYalValIsetrrpGlYtyrGlYCybaIagIn	220				
OY	675	GGAGGGGACGTGGCATACACCAATGTTTAAACAATATGTATCTCTGATAGGAACACT	734				
Db	221	ArgaentyrProglYvalTyrrThryValCybaenPhevalThrTripllegInserrThr	240				



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 8, 2003, 00:24:46 / Search time 72 Seconds  
(without alignments)  
6207.599 Million cell updates/sec

Title: US-10-037-270-482

Perfect score: 1560  
Sequence: 1 ggcactaccctccctagctca.....cgggaatgtagcatcactagt 866

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame+np.model -DEV=xlh  
-O=gsn2.1/USPTO\_spool/US10037270/runat\_07112003\_140515\_8759/app\_query.fasta\_1.1031  
-DB=SPTREMBL\_23 -QMT=faeacan -SUFFIX=rcpt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10037270 @CGCN 1.1 71 @runat\_07112003\_140515\_8759 -NCPUS=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: SP archaea:\*
- 2: SP bacteria:\*
- 3: SP fungi:\*
- 4: SP human:\*
- 5: SP invertebrate:\*
- 6: SP mammal:\*
- 7: SP mnc:\*
- 8: SP organelle:\*
- 9: SP phage:\*
- 10: SP plant:\*
- 11: SP rodent:\*
- 12: SP virus:\*
- 13: SP vertebrate:\*
- 14: SP unclassified:\*
- 15: SP tvirus:\*
- 16: SP bacteriaph:\*
- 17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1065	68.3	237 11 Q9DAA4	Q9dAA4 mus musculu

2	483	31.0	112 11	Q9CUH3	Q9cuH3 mus musculu
3	347.5	22.3	246 11	Q9ROT7	Q9roT7 mus musculu
4	347.5	22.3	251 4	Q8N2U3	Q8n2U3 homo sapien
5	341.5	21.9	246 11	Q9QUX9	Q9quX9 mus musculu
6	331	21.2	247 4	Q8NRM4	Q8nrm4 homo sapien
7	328.5	21.1	246 11	Q9ZIR9	Q9zir9 mus musculu
8	328.5	21.1	247 11	Q9D7Y7	Q9d7Y7 mus musculu
9	328.5	21.1	247 11	Q9CPN9	Q9cpN9 mus musculu
10	327.5	21.0	239 4	Q8N1C9	Q8n1C9 homo sapien
11	315	20.2	247 11	Q9CPN7	Q9cpN7 mus musculu
12	311	19.9	242 13	Q9W7Q7	Q9w7Q7 paralicthy
13	305.5	19.6	238 13	Q9W7Q6	Q9w7Q6 paralicthy
14	305	19.6	241 4	Q81YP2	Q81yp2 homo sapien
15	299.5	19.2	241 13	Q98TG9	Q98tG9 engraulis j
16	299.5	19.2	244 13	Q42159	Q42159 petromyzon
17	299	19.2	237 13	Q91515	Q91515 fugu rubrip
18	297.5	19.1	245 13	Q42160	Q42160 petromyzon
19	290.5	18.6	240 13	Q98TR0	Q98tr0 engraulis j
20	290	18.6	242 13	Q93266	Q93266 pseudopleur
21	287	18.4	242 13	Q92099	Q92099 parnotothe
22	284	18.2	241 11	Q8BW11	Q8bw11 mus musculu
23	282.5	18.1	247 13	Q42158	Q42158 petromyzon
24	282.5	18.1	247 13	Q42608	Q42608 petromyzon
25	282	18.1	222 13	Q8AV11	Q8av11 oncorhynch
26	280	17.9	255 4	Q96RQ0	Q96rq0 homo sapien
27	276.5	17.7	243 13	Q8AVB3	Q8avB3 brachydanto
28	267	17.1	219 13	Q91036	Q91036 gadus morhu
29	263	16.9	244 13	Q8QW3	Q8qw3 anguilla ja
30	261.5	16.8	251 11	Q9D9G7	Q9d9G7 mus musculu
31	260.5	16.7	251 11	Q54854	Q54854 ratu norv
32	255.5	16.4	253 11	Q91Y82	Q91y82 mus musculu
33	253.5	16.2	178 13	Q93594	Q93594 dicentrarch
34	253.5	16.2	250 13	Q93265	Q93265 pseudopleur
35	252.5	16.2	246 11	Q88301	Q88301 mus musculu
36	249	16.0	260 4	Q81W69	Q81w69 homo sapien
37	247	15.8	293 13	Q92046	Q92046 dissostichu
38	244	15.6	243 11	Q9D140	Q9d140 mus musculu
39	242	15.5	249 13	Q9W6K0	Q9w6K0 notochtha
40	242	15.5	249 13	Q8JFQ7	Q8jfQ7 gadus morhu
41	240	15.4	675 13	Q9W608	Q9w608 dissostichu
42	238.5	15.3	249 11	Q91VE3	Q91ve3 mus musculu
43	238	15.3	344 13	Q9W609	Q9w609 dissostichu
44	237	15.2	248 11	Q63224	Q63224 ratu norv
45	233	14.9	255 11	Q9JIS2	Q9jis2 mus musculu

ALIGNMENTS

RESULT 1	ID	Q9DAA4	PRELIMINARY:	PRT:	237 AA.
AC	Q9DAA4				
DT	01-JUN-2001	(TREMBlrel. 17, Created)			
DT	01-JUN-2001	(TREMBlrel. 17, Last sequence update)			
DT	01-MAR-2003	(TREMBlrel. 23, Last annotation update)			
DE	1700016G05R1k	protein (Hypothetical serine proteases).			
GN	1700016G05R1K.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Testis;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Aizawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamaneke S.,				
RA	Saito T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsumo Y., Nikaido I., Pesole G., Queckenbush J.,				
RA	Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,				



RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bersh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bulc C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Holman M., Hume D.A., Kamuya M., Lee N.H.,  
RA Lyons P., Marchionni L., Maahmud M., Mazzarelli V., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schenkebach C., Seya T., Shibaie Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Willeker C., Wlinding L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RT Nature 409:685-690(2001).  
RN [2]

SEQUENCE FROM N.A.  
RP STRAIN=C57BL/6J; TISSUE=Testis;  
RC MEDLINE=32354683; PubMed=12466851;  
RX  
RA THE FANTOM Consortium.  
RA THE RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RT Nature 420:563-573(2002).  
CC -  
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC EMBL: AK006028; BAB24373.1; -  
DR EMBL: AK002980; BAC26613.1; -  
DR HSSP: P00761; IAKS.  
DR MEROPS: S01.989; -  
DR MGD: MGI:1914940; 1700016G05R1K.  
DR Interpro: IPR001314; ChymoTrypsin.  
DR Interpro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYD\_SPC; 1.  
DR PROSITE: PSS0240; TRYPSIN\_DOM; 1.  
DR Hypothetical protein; Hydrolyase; Protease; Serine protease.  
KW Hypothetical protein; Hydrolyase; Protease; Serine protease.  
QO SSQUENCE 237 AA; 2669 MW; EF91396556B0394A CRC64;

Alignment Scores:	
Pred. No.:	2,06e-101
Score:	1065.00
Percent Similarity:	90.64%
Best Local Similarity:	83.40%
Query Match:	68.27%
DB:	11
US-10-037-270-482 (1-866) x Q9DAAA (1-237)	
	Length: 237
	Matches: 196
	Conservative: 17
	Mismatches: 22
	Indels: 0
	Gaps: 0

[illegible]

Qy	462	GAAGAAAGGAGCGGACACCCGTGATCTTGGGAGAAACCTGGAGGCCGCCGTGATGTGAT	521
Db	141	GLuSenAenGcYARgHieRphsPseuNrgGlnbNmSeGulwAlaPProValMeThrAsp	160
Qy	522	CGAGAAATGCCAAAAACAGAACAGGAAAAAGCCACAGGAATTCCTTATGTGGAATTT	581
Db	161	LYsAsPCyGlgLlyeThrGlnGlnGlnYseSerXhIaRgaSenSerLeuCyValaTrPhe	180
Qy	582	GTCGAAAGTATTCAGCGCGAATTTTGGGAGGCGCGTGTGTAATGTCATCTGCAAGAC	641
Db	181	ValLySvalPheSerNrgGlliePheGlyGulValaAlaAlaThrValIleGyGlyAsn	200
Qy	642	AACCTTCAGGGAATCGAGGTGGGACCTTCATTCGAGGGGACGTGCGCATCTACCAAT	701
Db	201	LYsLeuGlnGlyIleGluValGlyYHiePheMetGlyGlyAspValGlyIleTyTrhAsn	220
Qy	702	GTTTACAATATGTATCCGTATCTAGAAACCTGTAAGCAAG	746
Db	221	ILeTySerTyValaProTrpIleGluTySerThrThrTyGlyGly	235

RESULT 2	
09CUM3	
ID	09CUM3
AC	09CUM3
DT	01-JUN-2001 (TRENBLREL. 17, Created)
DT	01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT	01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE	1700016G05RIK protein (Fragment).
GN	1700016G05RIK.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;	

RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=10855660; PubMed=11217851;  
RA Kawai J, Shingawa A, Shibata K, Yoshino M, Itoh M, Ishii Y,  
RA Atakawa T, Hata A, Fukunishi Y, Konno H, Adachi U, Fukuda S,  
RA Saito K, Izawa M, Nishii K, Kiyosawa H, Kondo S, Yamataka I,  
RA Aizawa T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R,  
RA Kadota K, Matsuda H.A, Ashburner M, Batlov S, Casavant T,  
RA Fleischmann W, Gaasterland T, Gissi C, King B, Kochwa H,  
RA Kuehl P, Lewis S, Matsuo Y, Nikaado I, Pesole G, Quackenbush J,  
RA Schirml L.M, Staudli F, Suzuki R, Tomita M, Wagner L, Maehio T,  
RA Sakai K, Okido T, Furuno M, Kono H, Baldarelli R, Barch G,  
RA Blake J, Boffelli D, Bojunga N, Carninci P, de Bonaldo M.F.,  
RA Brownstein M.J., Bult C, Fletcher C, Fujita K, Gariboldi M.,  
RA Gustincich S, Hill D, Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U, Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontecki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC - - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL: AK016088; BAB30109.1; -  
DR HSSP: P00761; IAKS.  
DR MEROPS: S01.989; -  
DR MGD; MG1:1914940; 1700016G05RLK.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
KM Hydroxylase; Protease; Serine protease.  
FT NON\_TER 112  
SQ SEQUENCE 112 AA; 12764 MW; B95A276D81352923 CRC64;

Alignment Scores: 3.74e-41 length: 112  
Pred. No.:



```

QY 588 GTATTCAGCCGAATTTTGGGAGGTGGCCGTTGCTACTGTCAATCTGCAAGAACGCTC 647
Db 191 GtGlyLyAspSerCysGlnGlyAspSerGlyProValValCysAsnGlyGlnLeu 210
QY 648 CAGGGAATC-----GAGTGGGGCAGCTTCAATGAGGAGGAGCGTGGCATCTACACC 698
Db 211 GtGlyYlLeValSerTrpGlyTyGlyCysAlaLeuAspAsnProGlyValTyTrHr 230
QY 699 AATGTTACAAATATGATCTGATTCAGAACACT 734
Db 231 LysValCysAsnTyValAspTrpIleGlnAsnThr 242

RESULT 4
Q8N2U3 PRELIMINARY; PRT; 251 AA.
ID Q8N2U3;
AC 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypoetical protein (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;
RX [1]_SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strauberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC030238; AAH30238.1;
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
KM Hypoetical protein; Hydrolase; Protease; Serine protease.
FT NON_TER
SQ SEQUENCE 251 AA; 27091 MW; 53D3993B585AB328 CRC64;

Alignment Scores:
Pred. No.: 5 28e-27 Length: 251
Score: 347.50 Matches: 82
Percent Similarity: 52.59% Conservative: 40
Best Local Similarity: 35.34% Mismatches: 103
Query Match: 22.28% Indels: 7
Gaps: 3

US-10-037-270-482 (1-866) x Q8N2U3 (1-251)
QY 60 TTGGGTGCTCGCTGGAGCATTTTCTTGTGACTCATCTGTGAG----- 107
Db 15 LcGlyLThrValAlaValProPheAspAspAspAlaValLeValGlyTyTrHrCys 34
QY 108 AAGAAGACCTCTCTCTTTTGGTACTCAAGTCTCACTTCAACCCCTGTGGGC 167
Db 35 GtGlnAsnSerLeuProTyGlnValSerLeuAsnSerGlySerHisPheCysGlyGly 54
QY 168 GTCTCATCAACACCGAGTGGTGGGCGCCGCTCACTGATTTTACCAATTCGAAA 227
Db 55 SerLeuIleSerGlnGlnTrpValValSerAlaIleHisCysTyTrHrValGln 74
QY 228 GTGATCTGGGAATTTCAAGACAGAGTCAAGACGCTACTGACAGACGAATTAACCC 287
Db 75 ValArgLcGlyLysHisAsnIleLysValLeuGlnGlyAsnGlnGlnPheIleAsnAla 94
QY 288 ATTCAGATGTCGCGTACTGAGAACTACAGTCATAGGCGCCCAAGATGACCTCATGCTC 347
Db 95 AlaValIleLeuArgHisProLysTyAsnArgAspThrLeuAspAsnIleMetLeu 114

```

```

QY 348 ATCAAGCTGAGCTAAGCCGTCGATCTCAATCCCAAGTCACCCCTTCCTCGCAC 407
Db 115 IleLysLeuSerSerProAlaValIleAsnAlaArgValSerThrIleSerLeuProThr 134
QY 408 ACCAATGTACAGCCAGGACCTGTCTTACTTCACTTCAAGTTTGGATGAGCCCAAGAAC 467
Db 135 AlaProProAlaIleAlaGlyTrpGlnGlyCysLeuIleSerGlyTrpGlyAsnThrLeuSerPhe 154
QY 468 AGTGGCGGACACCCCTGACTTGGCGGAGAACCTGGAGCCCGCGTGAATGTCAGAA 527
Db 155 GlyAlaAspTyTrProAspGlnLeuLysCysLeuAspAlaProValLeuThrGlnAlaGlu 174
QY 528 TGCCAAAAACAGACACAGAAAGAAAGCCACAGAAATCTTATGTCGAATTTGCGAA 587
Db 175 CysLeValAlaSerTyTrProGlyLysIleThrAsnSerMetPheCysValGlyPheLeuGlu 194
QY 588 GTATTCAGCCGAATTTTGGGAGGTGGCCGTTGCTACTGTCAATCTGCAAGAACGCTC 647
Db 195 GtGlyLyAspSerCysGlnArgAspSerGlyGlyProValValCysAsnGlyGlnLeu 214
QY 648 CAGGGA---ATCAGAGTGGGGCAGCTTCAATGAGGAA-----GGGAGCGTGGCATCTACACC 698
Db 215 GtGlyValValSerTrpGlyHisGlyCysAlaTrpLysAsnArgProGlyValTyTrHr 234
QY 699 AATGTTACAAATATGATCTGATTCAGAACACT 734
Db 235 LysValTyAsnTyValAspTrpIleLeuAspThr 246

RESULT 5
Q9OUK9 PRELIMINARY; PRT; 246 AA.
ID Q9OUK9;
AC 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TESP4 (0910001B1PRK protein) (Trypsinogen 9).
GN TC OR 0910001B1PRK OR TRYPSINOGEN.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_Taxid=10090;
RX [1]_SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA MEDLINE=99436155; PubMed=10506205;
RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasakawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glasi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barch G.,
RA Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombærts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";

```

RN Nature0409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=21103195; PubMed=11160223;  
 RA Chen F., Rowen L., Hood L., Rothenberg E.V.;  
 RT "Differential transcriptional regulation of individual TCR Vbeta  
 segments before gene rearrangement.";  
 J. Immunol. 166:1771-1780(2001).  
 CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AB009661; BAA85187.1; -  
 DR EMBL; AB017031; BAA74760.1; -  
 DR EMBL; AK007843; BAB25300.1; -  
 DR EMBL; AE000664; BAB69057.1; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.058; -  
 DR MGD; MG1.1913350; 0910001B19RIK.  
 DR InterPro: IPR001314; ChymoTrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; Trypsin\_1.  
 DR PRINTS; PRO0722; CHYMOTRYP SIN.  
 DR SMART; SM00020; TRYD\_SPC; 1.  
 DR PROSITE; PSS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PSS0134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PSS00135; TRYP SIN\_SER; 1.  
 KM HydroLase; Protease; Serine protease.  
 QO SEQUENCE 246 AA; 26227 MW; 915C92353C42809 CRC64;

Alignment Scores:		
Pred. No.:	2, 2e-26	length: 246
Score:	341.50	Matches: 74
Percent Similarity:	54.72%	Conservative: 42
Best Local Similarity:	34.91%	Mismatches: 53
Query Match:	21.89%	Indels: 3
DB:	11	Gaps: 1

Qy	108	AAGAAGACCCTGCTCCCTATTGGGTACCTCAAGTCTCACTCAACCCCTGTGTGGC	167
Db	31	ArgGlnAsnSerIleProTyrGlnValSerLeuAsnSerGlyTyrHisPheCysGlyGln	50
Qy	168	GTCCTCATMAACCCAGCTGGGGGCGGCCAGCTCACTGTGATTTTACCAATCTGAAA	227
Db	51	SerLeuIleAsnAspGlnTrpValValSerAlaAlaHisCysTyrLysThrArgIleGln	70
Qy	228	GTGATGCTGGGAATTTCAAGACAGAGTCAAGACGGTACTGAAACAGACAATTAACCCC	287
Db	71	ValArgLeuGlnLysIleHisIleAsnIleAsnValLeuGlnGlnGlnPheValAsnSer	90
Qy	288	ATTCAAGATGTCGCCGCTACTGGAATTAACACTCATACGCCCCCAACAGATGACCTGCTC	347
Db	91	AlaIleIleLeuHisIleProAsnPheAsnSerArgThrLeuAsnAsnAspIleMetLeu	110
Qy	348	ATCAAGCTGGCTAAAGCTTGCATGCTCAATCCCAAGTCAAGCCCTTCCCTGCCACC	407
Db	111	IleIysLeuAlaSerProValThrLeuAsnAlaArgValAlaIleThrValAlaLeuProSer	130
Qy	408	ACCAATGTCAAGCCAGACGTCGTCTACTACTCAGAGTTTGACTGGACGCAAGAAAC	467
Db	131	SerCysAlaProAlaIleTyrGlnCysLeuIleSerGlyTTPGlyAsnThrLeuSerPhe	150
Qy	468	AGTGGCCGACACCCTGACTTTCGGCGAGAACTGGAAGGCCCCCGTGAATGTCATCGAAA	527
Db	151	GlyValAsnAsnProAsnLeuGlnGlnCysLeuAsnAlaProLeuLeuProGlnAlaAsp	170
Qy	528	TGCCAAAAAACAGAACAGAAAAACCCACAGAAATTCCTATGTGTGAATTTGTGAAA	587
Db	171	CysGlnAlaSerTyrProGlyLysIleThrAsnAsnMetIleCysValGlyPheLeuGln	190
Qy	588	GTAATTCAGCCGAATTTTGGGAGAGTGGCCGCTTGCTACTGTGCATCTGTGAAGAACAATC	647
Db	191	GlyGlyLysAsnSerCysGlnGlyAsnSerGlyGlyProValValCysAsnGlyGlnLeu	210

Qy	648	CAGGGAATC-----GAGSTGGGGCATTCATGGAGGGAGACGTCGGATCTACAC	699
Db	211	GlnGlyIleValSerTyrPheTyrGlyCysAlaIleuYsaAspAsnProGlyValThr	230
Qy	699	AATGTTACAAATATGATCTCCGATTTAGAACACT	734
Db	231	LysValCysAsnTyrValAspTyrIleGlnAspThr	242

RESULT	6
Q8NMH4	
ID	Q8NMH4
AC	Q8NMH4
DT	01-OCT-2002 (TREMBLrel. 22, Created)
DT	01-OCT-2002 (T-EMBLrel. 22, last sequence update)
DT	01-MAR-2003 (T-EMBLrel. 23, last annotation update)
DE	Trypsinogen C.
GN	TRY6.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=96256474; PubMed=8650574;
RX	Rosen L., Koop B.F., Hood L.;
RT	"The complete 685-kilobase DNA sequence of the human beta T cell
RL	receptor locus.";
SL	Science 272:1175-1162(1996).

RP SEQUENCE FROM N.A.  
 RA Rowen L., Seto J., Smilt A., Acharya C., Ahnarn M.E., Ankener M.,  
 RA Basbin D., Bungarner R., Chen L., Chen N., Deshpande P., Faust J.  
 RA Howard S., Jerome N., Koop B.F., Lee H., Loretz C., Paepser B.,  
 RA Zackrose K., Hood L.;  
 RT "Sequence determination of the human T cell receptor beta locus:  
 RT Strategy and error analysis";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
RA Rowen L., Wang K., Boyesen C., Ahearn M.E., Charnley P., Paepser B.,  
RA Lee I., Chen L., Trask B., Nickerson D., Soto D., Hood L.,  
RA "Sequence variation among several haplotypes in the human T cell  
RT receptor beta locus."  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBD databases.  
CC - Similarity: BELONGS TO THE PERLIDAB5 FAMILY S1.  
DR EMBL: U66061.1; AAC80208.1; -  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser. protease\_Try.  
DR Pfam: PF00089; trypsin. 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRY. SPEC. 1.  
DR PROSITE: PS00240; TRYPSIN\_DOM. 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS. 1.  
DR PROSITE: PS00135; TRYPSIN\_SER. 1.  
DR HydroLase; Protease; Serine Protease.  
KW SEQUENCE. 247 AA; 26539 MW; 6B9F58B6A4F757BE CRC64;

<b>Alignment Scores:</b>					
Pred. No.:	2,7e-25	Length:	247		
Score:	331.00	Matches:	78		
Percent Similarity:	50.43%	Conservative:	40		
Best Local Similarity:	33.33%	Mismatches:	108		
Query Match:	21.22%	Indels:	8		
DB:	4	Gaps:	2		
US-10-037-270-482 (1-866) x Q8NHM4 (1-247)					
OY	57 TATTGGGTCTCGTGGACATTTTTCGTAACATCTGTACG-----	107			
	:::::			:::	
Dd	9 PheValcIyAlaIalavalalavalProheAsphasphasparayrIlevalIgyltyr	28			
OY	108 -----AAGGAAGACCCGTCTCCCTATTTCGGTGTRACTCAAGTCACTTCAACCCTGT	161			

Db	29	ThrcyagelugluhsnserValProlyrGlnValSerleuAenSerglySerthierPheCys	48
Qy	162	GTGGGCGCTTCATCAAAACCCAGCTGGGTGGTGGCCCGCCAGCTACTGCTATTTCACAA	221
Db	49	GlyglYserleuilesergluIntPrValserlaGlyhnicGyetylyrProhis	68
Qy	222	CTGAAAGTGAATGCTGGGAAATTTCAAGACAGTCAGACAGCGTACTGAAACACACAA	281
Db	69	IlleGlnValArgheuglYglunihsaenileglValleuGlnGlyAvauglGlnPheIle	88
Qy	282	AACCCCATCAAGTCGCGCTACTGGAATCACTAGTCATAGCGCCCAACAGATGACTTC	341
Db	89	AsnAlaAlaValylleleArgHisrProlyrTyAsnArgIleThrLeuAsnAsnAspIle	108
Qy	342	ATGCTCATCAAGCTGGTAAAGCTGGCATGCTCAATCCCAAGTCCAGCCCTTCCTCC	401
Db	109	MetleuileuYserleuSerThrProAlaValIleAsnAlaHisValiserThrIleSerleu	128
Qy	402	GCCACCCACCAATGTCAGGCCAGGACACTGTCTGTCTACTCTCAGGTTTGACACTGAGCCAA	461
Db	129	ProthrIaerProAlaAlaGlyThrGluCysleuileSerglyTyrGlyAsnThrleu	148
Qy	462	GAAACAGCTGGCGGACACCCCTGACTTGGCGAGAACCTGGAGCGCCCGGTATGTCAT	521
Db	149	SerSerIlyAlaAspTyrrProAspGluLeuGlnCysleuAspAlaPravalLeuThrGln	168
Qy	522	CGAAGATCCCAAAAACAGAACAGGAAACGACAGGAATCTTATGTGTGAAATTT	581
Db	169	AlaIeyCysylsAlaserTyrrProleuylVilethnserlyMetPheCysValGlyPhe	188
Qy	582	GTCGAAGTATTCAGCCCAATTTTGGGAGAGTGGCCGTTGCTACTGTGATTCGAAAGAC	641
Db	189	LeuGlnGlyglYAspSerSergGlnGlyAspSerglyGlyProValIValCysAsnGly	208
Qy	642	AAGTCCAGGAGATC-----GAGGTGGGGCACTTATGAGGAGGAGAGTGGGATC	699
Db	209	GlnleuGlnIlyleValserTyrGlyCysAlaGlnLyAsnArgPrGlyVal	228
Qy	693	TACACCAATGTTTCAAAATATATATCTGATCTGATGAGAACACT	744
Db	229	TyrThrlyValIlyrAsnTyrrValAspTyrIlelyAspThr	242

RESULT 7

Q921R9 PRELIMINARY; PRT; 246 AA.

AC Q921R9;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DE 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)

DE Trypsinogen 16.

GN TRYGN16 OR TRYPSINOGEN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=Balb/c;

RA Rowen L., Hood L.,

RT "Comparison between strains Balb/C and 129 in a region of the mouse T cell receptor beta locus.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DDAJ databases.

RN 12

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/c;

RA MEDLINE=21103195, PubMed=11160223;

RA Chen F., Rowen L., Hood L., Rothenberg E.V.;

RT "Differential transcriptional regulation of individual TCR Vbeta segments before gene rearrangement.";

RL J. Immunol. 166:1771-1780(2001).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

DR EMBL; AF107342; AAC79093.1; -.

[illegible]

DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 2210010C04Rik protein.  
 GN 2210010C04Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi U., Fukuda S.,  
 RA Saito K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Salto T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Salto R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Glaser C., King B., Kochiya H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boeckelli D., Boljuna N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
 RA Hayashizaki Y.:  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AK008695; BAB25837.1; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.151; -  
 DR WGD; MG11914623; 2210010C04Rik.  
 DR Interpro; IPRO01314; Chymotrypsin.  
 DR Interpro; IPRO01254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PSS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Protease; Serine protease.  
 SO SEQUENCE 247 AA; 26407 MW; 84D474DB32B3A55 CRC64;

**RESULT 9**

Df	HspProAntyRnaAlaAnThrTyrAmbsnApIImeLeuIIeUlysleUlystrH	111
Oy	363 CCGCCATGCTCAATCCCAAGTCACGCCCTTCCCCTTGCCACCAACAATGCAGGCCA	422
Dd	117 AlAlatThrLeuAnsenserArgValserThrValAlalaLeuProArGerCyseProSerAla	136
Oy	423 GGCACTGTCTCTACTACTCAGGTTTGACATGGACCACCAAGAAGTCGCCGACACCT	482
Dd	137 GLythrArgCyseLvalValserScLyTrpLyAnInThrLeuSerSeScLyThraAnTyPro	156
Oy	483 GACTTGGCGGCAAGCATCTGAGAGCCCCCGCTGATGTGATCGAAGATGCCAAAAACAGAA	542
Dd	157 SerleuLeuGlncYcSleuAsPalaProValLeuSerAsPseSerCythrSerSerTyx	176
Oy	543 CAAGCAAAAAGCCACAGGAATTCTTATGTGTGAATTGTGAAGATTAATGAGCCGAATT	602
Dd	177 ProGlytysylIethrSerAsMetheCyseLeuGlncYcPhelLeuGlncYcLyAsPseSer	196
Oy	603 TTTGGGAGATGCGCCGCTTGCTACTACTCTGCAACCAAGCAAGCTCCAGGATC-----	656
Dd	197 CyseGlncYcAsPseSerGlyLyProValValCyseAngLyGlnLeuGlncYcLvalValSer	216
Oy	657 ---GAGNGGGGCGCACTTCATGGAGGGGAGCGCGGCACTTACCAACAATTTTCAATAAT	713
Dd	217 TrpGlyTyrgLyCyseAlaGlnArgLyProGlyValTyThrLyValCyseUtyr	236
Oy	714 GTATCTTGATTTGAGAACACT 734	
Dd	237 ValAnTrpIleGlnGlnThr 243	
ID	OQCPN9 PRELIMINARY; PRT; 247 AA.	
AC	OQCPN9;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	
DE	2210010CD4RIK protein (Trypsinogen 7).	
GN	Mus musculus (Mouse).	
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TRISUB=Small intestine, and Pancreas;	
RX	MEDLINE=21085560; PubMed=11217851;	
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,	
RA	Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,	
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,	
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasuikawa T., Saito R.,	
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,	
RA	Frieschmann W., Gaasterland T., Gissi C., King B., Kochiya H.,	
RA	Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,	
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,	
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,	
RA	Blake J., Boffelli D., Bojungta N., Carninci P., de Bonaldo M.F.,	
RA	Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,	
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,	
RA	Lyonis P., Matchonni L., Mashima U., Mazzarelli I., Sakamoto P.,	
RA	Nordone P., Ring B., Ringueld M., Rodriguez I., Sakamoto P.,	
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shihata Y., Storch K.-F.,	
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitteker C., Wilming L.,	
RA	Wynshaw-Borrie A., Yoshida K., Haegawa Y., Kawaji H., Kohneut S.,	
RA	Hayashizaki Y.,	
RT	"Functional annotation of a full-length mouse cDNA collection.";	
RL	Nature 409:665-690(2001).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BALB/c;	
RX	MEDLINE=21103195; PubMed=11160223;	
RA	Chen F., Rowen L., Hood L., Rothenberg E.V.,	
RT	"Differential transcriptional regulation of individual TCR beta	



RT segments before gene rearrangement.";  
RL J. Immunol. 166:1771-1780(2001).  
DR EMBL; AK008460; BAB25680.1; -  
DR EMBL; AK007773; BAB25246.1; -  
DR EMBL; AE000663; AAB69044.1; -  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.151; -  
DR MGD; MGI:1914623; 2210010C04Rik.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR SMART; SM00020; TRY. SPC; 1.  
DR PROSITE; PSS0240; TRYPsin DOM; 1.  
DR PROSITE; PS00134; TRYPsin HIS; 1.  
DR PROSITE; PS00135; TRYPsin SER; 1.  
KM Hydroxylase; Protease; Serine protease.  
SQ SEQUENCE 247 AA; 26422 MW; B8C5767B182D9AAD CRC64;

## Alignment Scores:

Pred. No.:	4.9e-25	Length:	247
Score:	328.50	Matches:	75
Percent Similarity:	53.14%	Conservative:	35
Best Local Similarity:	36.23%	Mismatches:	94
Query Match:	21.06%	Indels:	3
DB:	11	Gaps:	1

US-10-037-270-482 (1-866) x Q9CPN9 (1-247)

QY 123 CCCGATTTGGTGTACCAAGTCTCACTTCAACCCCTGTGGGCGTCTATCAACCC 182  
DB 37 PCTGTGlnValSerLeuAsnSerGlyTrHisPheCysGlySerLeuIleAsnSer 56  
QY 183 AGCTGGTGTGGGCCCAAGTCTCACTTATTCACCAATGTGAAGATGTGGGAAT 242  
DB 57 GlnTrpValValSerAlaAlaHisCysTrpLysSerArgIleGlnValAglLeuGlyGlu 76  
QY 243 TTCAGAGCAGAGTCAGAGCGGTACTGAACAGCAATTAACCCATTCAGATGTCGCG 302  
DB 77 HisAsnIleAspAlaLeuGlnGlyGlyGlnPheIleAspAlaAlaLysIleIleArg 96  
QY 303 TACGTGACATCAAGTCTACCGCCCAAGATGACCTGATGTCATCAAGCGGTGAAG 362  
DB 97 HisProAsnTrpAsnAlaAsnThrTrpAsnAsnAspIleMetLeuIleLysLeuLysThr 116  
QY 363 CCTGGCATGCTCAATCCCAAGTCCAGCCCTTCCTCCGACACCAATGTGAGCCA 422  
DB 117 AlaAlaHisLeuAsnSerArgValSerThrValAlaLeuProArgSerCysProSerAla 136  
QY 423 GGCATGCTGTCTACTCTCAAGTTTGAGTGAAGCAAGAAACAGTGGCCGACACCT 482  
DB 137 GlyThrArgCysLeuValSerGlyTrpGlyAsnThrLeuSerSerGlyThrAsnTrpPro 156  
QY 483 GACTGGCGGAGAACCTGGAGCGCCCGGTGATGTCTGATGAGAAATCCAAAAACGAA 542  
DB 157 SerLeuLeuIleCysLeuAspAlaProValLeuSerAspSerCysThrSerSerTyr 176  
QY 543 CAAGAAAGAAAGCAAGAGATTCCTTATGTGTGAATTTGAAAGATTCAGCCGAAAT 602  
DB 177 ProGlyLysIleThrSerAsnMetPheCysLeuGlyPheLeuGlnGlyLysAspSer 186  
QY 603 TTTGGGAGGTGGCGCTGTCTACTGTGATGTCAAGAGCAAGTCCAGGGAATC----- 656  
DB 197 CysGlnIleLysAspSerGlyLysProValIleCysAsnGlyLysLeuGlnGlyValIleSer 216  
QY 657 ---GAGTGGGCGCATTCATGAGGAGGAGCGTGGCATTCACCAATGTTTACAAATAT 713  
DB 217 TrpGlyLysIleCysAlaGlnArgGlyLysProGlyValIleThrLysValCysLysTrp 236  
QY 714 GTATCTGATGATGAGACACT 734  
DB 237 ValAsnTrpIleGlnGlnThr 243

RESULT 10  
Q8NIC9

ID Q8NIC9 PRELIMINARY; PRT; 239 AA.  
AC Q8NIC9;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Protease, serine, 2 (trypsin 2).  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RA Strausberg R.;  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; BC030260; AAH30260.1; -  
DR InterPro; IPR001314; Chymotrypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRY. SPC; 1.  
DR PROSITE; PSS0240; TRYPsin DOM; 1.  
DR PROSITE; PS00134; TRYPsin HIS; 1.  
DR PROSITE; PS00135; TRYPsin SER; 1.  
KM Hydroxylase; Protease; Serine protease.  
SQ SEQUENCE 239 AA; 25687 MW; 1389BCE180DBED CRC64;

## Alignment Scores:

Pred. No.:	6.17e-25	Length:	239
Score:	327.50	Matches:	74
Percent Similarity:	52.15%	Conservative:	35
Best Local Similarity:	35.41%	Mismatches:	97
Query Match:	20.99%	Indels:	3
DB:	4	Gaps:	1

US-10-037-270-482 (1-866) x Q8NIC9 (1-239)

QY 108 AAGAGAGCCCTGCTCCTATTGTTGGTGAAGTCTCACTTCAACCCCTGTGGGC 167  
DB 31 GlnGlnAsnSerValProTrpGlnValSerLeuAsnSerGlyTrHisPheCysGlyGly 50  
QY 168 GTCTCATCAAAACCAAGTGGTGGTCTGGCCCAAGTCACTGTTATTCACCAATGAAA 227  
DB 51 SerLeuIleSerCysGlnTrpValValSerAlaGlyHisCysTrpLysSerArgIleGln 70  
QY 228 GTGATGTGGGAAATTCAGAGCAGATCAGAGCGTACTGAACAGCAATTAACCCC 287  
DB 71 ValArgLeuGlyGlnHisAsnIleGlnValLeuGlnGlyAsnGlnPheIleAsnAla 90  
QY 288 ATTCAATGTGCTGGTCAAGTCAAGTCAATGAGCCCAAGATGACCTGATGCTC 347  
DB 91 AlaLysIleIleLeuArgHisProLysTrpAsnSerArgThrLeuAspAsnAspIleLeuLeu 110  
QY 348 ATCAAGCTGATGAAGCTGATGCTCAATCCCAAGTCCAGCCCTTCCTCCGACAC 407  
DB 111 IleLysLeuSerSerProAlaValIleAsnSerArgValSerAlaIleSerLeuProThr 130  
QY 408 ACCAATGTCAAGCCAGGACACTGTCTGTACTCTCAAGTTTGAGCTGAGCAAGAAAC 467  
DB 131 AlaProProAlaIleGlyIleGlnSerLeuIleSerGlyTrpGlyAsnThrLeuSerSer 150  
QY 468 AGTGGCGGACACCTGATCTTGGCGGAGAACCTGGAGCGCCCGTGTATGTCTGATGAGA 527  
DB 151 GlyAlaAspTrpProAspGlnLeuGlnCysLeuAspAlaProValLeuSerGlnAlaGln 170  
QY 528 TGCCAAAGAAAGCAAGAGAAAGCAAGAGATTCCTTATGTGTGAATTTGAAA 587  
DB 171 CysGlnAlaSerTrpProGlyLysIleThrAsnAsnMetPheCysValGlyPheLeuGln 190  
QY 588 GTATTCAGCGGAATTTTGGGAGGTGGCGCTGTGATCTGATCTGCAAGCAAGCTC 647  
DB 191 GlyGlyLysAspSerCysGlnGlnLysAspSerGlyLysProValIleSerAsnGlnGlnLeu 210



QY 648 CAGGGAATC-----GAGTGGGCACTTCATGGAGGAGCGGCACTGCATTCACCC 698  
|||  
Db 211 GInGlyIleValserTrrpGlyTyrglyCyAlaGlnlyAsnlyMrProGlyValTyThr 230  
|||  
QY 699 AATGTTTACAAATATGATCTCGATT 725  
|||  
Db 231 LyValTyAsnTyValAspTrrpIle 239  
|||  
RESULT 11  
Q9CPN7 PRELIMINARY, PRT, 247 AA.  
ID Q9CPN7  
AC Q9CPN7  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE 1810009J06RIK protein (TRYPSINogen 4).  
GN 1810009J06RIK OR TRYPSINOGEN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Pancreas;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hata A., Fukunishi Y., Komio H., Adachi U., Fukuda S.,  
RA Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirni L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guncionich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombert P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,  
RA Wyszewski B., Yoshida K., Hasegawa Y., Kawai H., Kohesuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE=21103195; PubMed=11160223;  
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;  
RT "Differential transcriptional regulation of individual TCR beta  
RT segments before gene rearrangement.";  
RL J. Immunol. 166:1771-1780(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL: AK007406; BAB25018.1; -;  
DR EMBL: A5000663; AAB69055.1; -;  
DR HSSP: P00763; IDPO.  
DR MEROPS: S01.129; -;  
DR MGD: MGI:1920876; 1810009J06RIK.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin\_1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP\_SPC\_1.  
DR PROSITE: PS00240; TRYPSIN\_DOM\_1.  
DR PROSITE: PS00134; TRYPSIN\_HIS\_1.  
DR PROSITE: PS00135; TRYPSIN\_SER\_1.  
KM Hydrolyase; Protease; Serine protease.  
SQ SEQUENCE 247 AA; 26503 MW; EDF5F7696833C7BC CRC64;

Alignment Scores: 1.23e-23 Length: 247  
Pred. No.:

Score: 315.00 Matches: 81  
Percent Similarity: 53.27% Conservative: 33  
Best Local Similarity: 37.85% Mismatches: 84  
Query Match: 20.19% Indels: 16  
DB: 11 Gaps: 6  
US-10-037-270-482 (1-866) x Q9CPN7 (1-247)  
QY 123 CCCTATTGGTGTACCTCAAG-----TGTCACTCAACCCCTGTGGCGTCTC 173  
|||  
Db 36 TrpTyGlnValSerLeuAsnAspGlyIleSerHis-----GInGlySerLeu 53  
|||  
QY 174 ATCAAAACCCAGCTGGTGGTGGCCCACTCACTGATTTTCAAAATGAAAGATG 233  
|||  
Db 54 TleSerAspGlnTrpValLeuSerAlaAlaHisCyTyLyValArgLeuGlnValArg 73  
|||  
QY 234 CTGGGAATTTCAAGACAGACAGACTGACAGACGGTACTGAAACAGCAATTAACCCATTG 293  
|||  
Db 74 LeuGlyGlnHisValAsnIleAspValLeuGlyGlyGlnGlnPheIleAspAlaGlyLeu 93  
|||  
QY 294 ATGTGCGGTACTGGAACATGACATGATACAGCCCAAGATGACCTCATCTCATCAAG 353  
|||  
Db 94 TleIleArgHisProAspTyrrAsnTyrrAsnTyrrValAspAsnAspIleMetLeuIleLeu 113  
|||  
QY 354 CTGGCTTAACCTGCTCAATCTCAATCCCAAGTCCAG-----CCCTTCCCTC 401  
|||  
Db 114 LeuValSerProAlaIleLeuAsnSerGlnValSerThrValSerLeuProArgSerCy 133  
|||  
QY 402 GCCACCAACCAATGTCAAGCCAGCAGCACTGTCTTACTCTCAAGTTTGACCTGACCA 461  
|||  
Db 134 AlaSerThrAsnAlaGln-----CyLeuValSerGlyTyrrGlyAsnThrVal 149  
|||  
QY 462 GAAACACGCGCCGACACCTGACTGCGGCAAGCTGAGAGCCGCCCTGATGTGAT 521  
|||  
Db 150 SerIleGlyLyLeuTyrrProAlaLeuGlnGlnCyLeuGlnAlaProValLeuSerAla 169  
|||  
QY 522 CGAGATGCGCAAAACACACACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 581  
|||  
Db 170 SerSerCyLeuValSerTyrrProGlyGlnIleThrSerAsnMetCysLeuGlyPhe 189  
|||  
QY 582 GTGAAGTATTCAGCCGATTTTGGGAGGTGGCCGTGCTACTGATCTGCAAAAGAC 641  
|||  
Db 190 LeuGlnGlyGlyLyAspSerCyAspAspGlyAspSerGlyGlyProValValCyAsnGly 209  
|||  
QY 642 AAGCTCCAGGGAATC---GAGTGGGCACTTC-----ATGAGAGGGAAGCTGCGCAT 692  
|||  
Db 210 GInIleGlnGlyIleValserTrrpGlySerValCyAlaMetCysArgGlyLyAspProGlyVal 229  
|||  
QY 693 TACACCAATGTTTACAAATATGATCTCGATTGGAACACT 734  
|||  
Db 230 TyrrIleValValCyAsnTyrrLeuSerTrrpIleGlnGlyThr 243  
|||  
RESULT 12  
Q9W707 PRELIMINARY, PRT, 242 AA.  
ID Q9W707  
AC Q9W707  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Trypsinogen 1.  
OS Paraliichthyidae (Flounder).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
OC Pleuronectoidae; Paraliichthyidae; Paraliichthyae.  
OX NCBI\_TaxID=8255;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RA Suzuki T., Sliwastrava A.S., Kurokawa T.;  
RT "Japanese flounder mRNA for trypsinogen 1.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB029750; BAA62362.1; -;



```

Db      208  TrpGlyTYrGlyCysalacInargSerHisProGlyValTYrAlaIleValCysIleSph 22
QY      714  GRTCTCGATGATTGAGACACT 734
      ::: |||::||| |||
Db      228  IleasptRleungInargThr 234

RESULT 14
Q81YP2
ID      Q81YP2;          PRELIMINARY;          PRT;          241 AA.
AC      Q81YP2;
DT      01-MAR-2003 (T-EMBLrel. 23, Created)
DT      01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE      Hypochemical protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RL      Strassberg R.;
RL      Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
EMBL; BC035384; AAH35384.1; -.
KW      Hypochemical protein.
SQ      SEQUENCE 241 AA; 27084 MW; 964829FE2BAD60A CRC64;

```

[illegible]

QY	558	---AGGAATTCCTTATGTGTGTAATTTGGAAAGTATTCAGCGCAATTTTGGGAGAGTG	614
Db	175	ThrgluammetleCysvalGlyIleValIProGlyIaYgArgGlnProCysIuysGluVal	194
QY	615	GCCGCTGTACTGTCTATCTGCACAAACAGACTCCAGGAGATC-----GAGGTGGGG	665
		:::	
Db	195	SerIlaIaIaProAlaIleIleCysabamGlymetIeugInGlyIleIleuSerPheAlaIaGlyI	214
QY	666	CAC TTCATGAGGAGGGAGAGCTGGCGCATTCACCAATGTTACCAATATGATTCCTGGATT	725
		:::	
Db	215	CysValIleuAraGAlaAspValGlyIleTyrAlaValIlePheTyrTyrIleProIlePile	234
QY	726	GAGAACACTGCTTAAGAC	743
Db	235	GIuAsnValIleGlnAsn	240

ID	Q98TGS	PRELIMINARY;	PRT;	241 AA.
AC	Q98TGS9			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Trypsinogen II.			
GN	ATRYII.			
OS	Engraulis japonicus (Japanese anchovy).			
OC	Eukaryota; Metazoa; Chordata; Cratiacea; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;			
OX	NCBI_TaxID=42892;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pyloric caeca;			
RA	Watabe S., Ahsan M.N., Funabara D.;			
RT	"Anchovy trypsinogen mRNA.";			
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	EMBL; AB041930; BAB40330.1; ..			
DR	HSSP; P00763; IDPO.			
DR	MEROPS; S01.258; ..			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Ser_protease_Try.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; TRY_Spc; 1.			
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Protease; Serine protease.			
SO	SEQUENCE	241 AA;	26282 MW;	FE362039CAEEB2F6 CR664;

Fragment Sizes:			
Prod. No.:	4.95e-22	Length:	241
Score:	299.50	Matches:	75
Percent Similarity:	49.38%	Conservative:	45
Best Local Similarity:	30.86%	Mismatches:	104
Query Match:	19.20%	Indels:	19
DB:	13	Gaps:	4
US-10-037-270-482 (1-866) x Q98RG9 (1-241)			
Oy	42 ATGAATATGCTTCTATTTGGAGTGCCCGCTGGACATTTCCTTGCTGCATCATC	101	
Dd	1 MetargserleValaPheleuValleueuglyAlaIaPhalaIglaImprasyile	20	
Oy	102 GTTCAGAAAGAAGACCCTGTCCCTAT-----TTGCTGTACTCAAGTCTCAC	149	
Dd	21 ValIglylTyrgIucyGlnProtyrSerGlnProHISgInvalSerleuInseryIy	40	
Oy	150 TTCAACCCCTTGTGGGGCTCTATCAAACCCACTGGCTCTGGCCCCACTCATTCG	209	
Dd	41 TyrlAspheCyGlglySerleuValSerAspSerTrpValValSerIalalAhIsCyS	60	

```
QY 210 TATTACCAATCTGAAAGTATGCTGGAAATTTCAAGACAGAGTCAGAGACGGTACT 269
Db 61 TyrIyserarValGIuValArgMetGlyLuhIshisrIleGlyMetThrGluGlyAsn 80
QY 270 GAACAGACATTTAACCCATTACAGATGCTCCGCTACTGAACTACAGTCAATAGCCCCCA 329
Db 81 GluGlnhethIleAspSerSerArgValIleArgHisProGlnTyrAspSerTyrAsnIle 100
QY 330 CAGATGACCTCATGCTCATGCTAGCTAGCTGCTAGCTGCTCAATCCCAAGTCCAG 389
Db 101 AspAsnAspIleMetLeuIleLysLeuSerLysProIleThrLeuAsnGlnTyrValGln 120
QY 390 CCGCTTCCTCCGACACCAATGTCAGGCCAGGCACTGTCTGTACTCTCAGGTTTG 449
Db 121 ThrValAlaLeuProSerSerCysAlaProIleGlyThrMetCysLeuValSerGly--- 139
QY 450 GACTGAGCCCAAGAAACAGTGGCCGACACCTGACTTGGCGGCAAACTGGAGGCCCC 509
Db 140 ---TrrpGlyAsnThrMetSerAsnValSerGlyAspLysLeuGlnCysLeuGlnIlePro 158
QY 510 GTGATGCTGATCGAAGATGCAAAACAGAACCAAGAAAGCCACAGGAATTCCCTTA 569
Db 159 IleuSerAspAspGlyAspCysLysAsnSerTyrProGlyMetIleThrGluSerMetPhe 178
QY 570 TGTGTGAATTTGTGAAGATATTACGCCGAATTTTGGGAGGTGGCCGTTGCTACTGTC 629
Db 179 CysAlaGlyTyrLeuGlnGlyGlyLysAspSerCysGlnGlnLysAspSerGlyProVal 198
QY 630 ATCTGCAAGACAGCTCCAGGGATC-----GAGGTGGGG 665
Db 199 ValCysAsnGlyGluLeuGlnGlyIleValSerTrrpGlyTyrGlyCysAlaGluArgAsp 218
QY 666 CACTTCATGGAGGGGAGCTCGGCATACCAATGTTTACAATATGATCCTGGATT 725
Db 219 His-----ProGlyValTyrAlaLysValCysLeuPheAsnAspTrrpIle 233
QY 726 GAGAACACT 734
Db 234 AspSerThr 236
```

Search completed: November 8, 2003, 02:11:37  
Job time : 81 secs